UNIVERSITY OF CALIFORNIA, LOS ANGELES **UCLA**

BERKELEY • DAVIS • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

Timothy F. Malloy
Professor of Law
(310) 794-5278
Email: malloy@law.ucla.edu

SCHOOL OF LAW
405 HILGARD AVENUE
LOS ANGELES, CA 90095-1476

August 16, 2011

Debbie Raphael
Dept of Toxic Substances Control / CalEPA
1001 I Street, 25th Floor
Sacramento, CA 95814

Re: Stakeholder Elicitation for Alternatives Analysis

Debbie
Dear ~~Ms. Raphael~~:

I am writing on behalf of our research team to thank you for volunteering your time and experience as part of our research regarding alternatives analysis in the regulatory context. Your participation provided us with insights well beyond the data regarding the relative weights accorded to the various criteria, and we are very thankful that you were willing to devote so much time to the interviewing process.

I am enclosing a copy of the final project report we recently submitted to the Public Health Trust. While it is final with respect to closing out that grant, we are actively reviewing and revising the report for publication and dissemination more broadly. Should you be interested in reviewing the report, we would likewise be very grateful for any comments, criticisms or suggestions you might have.

Thanks again.

Sincerely,

Tim

Timothy F. Malloy
Professor of Law

Enclosure

UCLA Sustainable Technology & Policy Program
Public Health Trust Grant No. 1016356

Developing Regulatory Alternatives Analysis Methodologies for the California Green Chemistry Initiative

Final Report

August 1, 2011

stpp

Linking Law, Science and Policy to Spur Innovation

Sustainable Technology and Policy Program

The Sustainable Technology and Policy Program, a joint undertaking of the UCLA Schools of Law and Public Health, is an interdisciplinary research and education program. Its mission is supporting the development of effective, balanced chemical policies, and the spread of safer chemicals and alternative manufacturing processes in the marketplace. STPP brings together researchers from across the UCLA campus and beyond with non-governmental agencies, policymakers and businesses in a unique, action-oriented initiative.

STPP is pleased to acknowledge the generous support of the Public Health Trust for this project. Funding was provided by the Public Health Trust, a program of the Public Health Institute, through a defendant's settlement of a complaint brought by the State of California.

Timothy F. Malloy, J.D., Faculty Director, STPP; Professor of Law, UCLA School of Law
Peter J. Sinsheimer, Ph.D., MPH, Executive Director, STPP
Ann Blake, Ph.D., Environmental and Public Health Consulting
Igor Linkov, Ph.D., Adjunct Professor of Engineering and Public Policy, Carnegie Mellon University and Research Scientist, US Army Engineer Research and Development Center

The authors would like to acknowledge the research and analytical support of Carrie Gibson and Elina Nasser, which was central to the success of this project. The authors also acknowledge the support of the DECERNS team (lead by Professor Boris Yatsalo) for allowing the use of the DECERNS software.

Sustainable Technology and Policy Program
University of California, Los Angeles
650 Charles E. Young Drive, South
21-293 CHS
Los Angeles, CA 90095

www.stpp.ucla.edu

Table of Contents

List of Figures

List of Tables

List of Appendices

I. Introduction

A. Project Context

Chemicals in commerce is an issue that sorely needs attention. Each day people are exposed to thousands of chemicals, some of which are known to be hazardous and many of which have never been tested for their health or environmental effects. Potentially hazardous chemicals not only pose direct threat to public, but also could result in significant cost to industry through potential future liabilities. Over the last ten years, reform efforts in the United States, Canada and Europe have begun to find traction. Most famously, the European Chemicals Agency is currently implementing REACH, a comprehensive restructuring of chemical regulation in the European Community.[1] In the United States, the well-documented failure of the federal Toxic Substances Control Act (TSCA) to meaningfully regulate chemicals has led to reform efforts at the state and federal level.[2] Most notable among the state efforts is California's Assembly Bill 1879 (AB 1879), signed into law in 2008. Considered revolutionary by some, the California law was intended to overcome the failings of TSCA and usher in a new "green chemistry" approach focused on designing safer chemicals and products.

AB 1879 makes *alternatives analysis* the centerpiece of a new comprehensive regulatory program designed to reduce or eliminate significant chemical hazards in consumer products. The basic concept underlying AB 1879 is straightforward: manufacturers of commercial and consumer products ought to design safety into those products. In doing so, however, they should avoid "regrettable substitution," the replacement of one hazardous chemical with another presenting similar or even worse hazards. Alternatives analysis—the identification, assessment and comparative evaluation of alternatives to hazardous chemicals—can provide a transparent, rigorous methodology for avoiding regrettable substitution. Under that statute, once the California Department of Toxic Substances Control (DTSC) prioritizes a chemical of concern in a consumer product, an alternatives analysis must be performed. On the basis of the alternatives analysis, DTSC is charged with taking regulatory responses ranging from a ban of the chemical to product labeling.

Despite the central role that alternative analysis plays in AB 1879, the methodology itself is neither well-developed nor tailored to application in regulatory settings. The statute itself provides little in the way of guidance. To a limited degree, businesses, researchers, and non-governmental organizations have developed and implemented various forms of alternatives analysis.[3] However, these efforts were not undertaken in a regulatory context in which the

---

[1] Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:396:0001:0849:EN:PDF.

[2] Child, Worker, and Consumer-Safe Chemicals Act of 2005, S. 1391, 109th Cong. (2005); Child, Worker, and Consumer-Safe Chemicals Act of 2005, H.R. 4308, 109th Cong. (2005); Kid-Safe Chemicals Act of 2005, S. 3040, 110th Cong. (2008); Child, Worker, and Consumer-Safe Chemicals Act of 2005, H.R. 6100, 110th Cong. (2008) ("KSCA"); Safe Chemicals Act of 2011, S. 487, 112th Cong., (2011)

[3] *See* Brandon Kuczenski and Roland Geyer, CHEMICAL ALTERNATIVES ANALYSIS: MODELS, METHODS AND TOOL (August 2010); Peter J. Sinsheimer, *et al.*, *.The Viability of Professional Wet Cleaning as a Pollution Prevention Alternative to Perchloroethylene Dry Cleaning*, 57 Journal of the Air & Waste Management Association 172 (2007). Also see Toxics Use Reduction Institute, FIVE CHEMICALS ALTERNATIVES ASSESSMENT STUDY (July 2006); Timothy F. Malloy and Peter Sinsheimer, *Innovation, Regulation, and the Selection Environment*, 57 Rutgers L. Rev. 183 (2004).

methodology must be applied consistently and transparently across a range of settings.[4] In the regulatory setting, agencies have engaged in the comparison of chemical alternatives to limited degrees. For example, under the federal Toxic Substance Control Act, EPA engaged in an alternatives analysis relating to asbestos-containing products as part of its ill-fated rulemaking in the late 1980's.[5] That rule was overturned by a federal court, in part based upon deficiencies in EPA's alternatives analysis methodology.[6] More recently, EPA's Significant New Alternatives Program (SNAP) uses a relatively simple method to verify the safety of substitutes for ozone-depleting compounds, relying upon a series of heuristic guiding principles coupled with a qualitative comparative risk approach.[7] The European Chemical Agency has also generated guidance for alternative analysis of certain regulated chemicals under REACH, but those methods stop short of providing an integrated decision platform and have not yet been widely implemented.[8]

The design and deployment of alternatives analysis has been a major focus of rulemaking under AB 1879, as DTSC and stakeholders from varied perspectives struggle with both broad concepts and finer grained details. AB 1879 lays out criteria and metrics for alternative analysis, which makes the first step towards comprehensive evaluation framework. A well-designed alternative analysis methodology would clearly advance the central goals of AB 1879, and represent a fundamental improvement in the protection of public health and the environment. On the other hand, however, a poorly conceived methodology could mire the regulatory program in the same problems that often cripple the conventional risk management approach currently used for chemical regulation.

This pilot project develops and evaluates an alternatives analysis methodology and supporting decision-analysis software for use in a regulatory context. The project applied the methodology and software to two case studies to illustrate the feasibility of their application in regulatory settings and to examine the potential value and limitations of this type of approach to alternatives analysis.

B. Project Overview

Alternatives analysis is a scientific method for prioritizing different courses of action; in this case for determining the viability of safer substitutes for existing products or processes that use hazardous substances. Alternatives may include drop-in chemical substitutes, material substitutes, changes to manufacturing operations, and changes to component/product design.[9] The methodology compares the alternatives to the regulated product and to one another across a variety of attributes, typically including public health, environmental, technical and economic.

---

[4] Lars Koch and Nicholas Ashford, *Rethinking the Role of Information in Chemicals Policy: Implications for TSCA and REACH*, 14 Journal of Cleaner Production 31 (2006).
[5] 51 Fed. Reg. 3738 (1986); ICF Incorporated, Regulatory Impact Analysis of Controls on Asbestos and Asbestos Products, Appendix H (January 19, 1989).
[6] Corrosion Proof Fittings, *et al.* v. EPA, 947 F.2d 1201 (5th Cir. 1991).
[7] See 59 Fed. Reg. at 13044, 13046 (March 18, 1994).
[8] *See* ECHA, Guidance on the Preparation of an Application for Authorisation (ECHA-11-G-01, January 2011); ECHA, Guidance for the Preparation of an Annex XV Dossier for Restrictions (June 2007).
[9] Peter J. Sinsheimer, *et al.*, *The Viability of Professional Wet Cleaning as a Pollution Prevention Alternative to Perchloroethylene Dry Cleaning*, 57 J. AWMA 172 (2007).

Alternatives analysis consists of two separate yet related components, as shown in Figure I-1.[10] The first, *alternatives assessment*, includes (1) identification of the key criteria for comparison (*e.g.,* technical, health and safety, environmental, and economic attributes) and the metrics for measuring performance on those criteria, (2) identification of potentially viable alternatives, and (3) collection and compilation of data regarding performance of the regulated product and alternatives for each criterion. While alternatives assessment is largely a data-driven, objective process, it involves significant application of best professional judgment and discretion. The second component of alternatives analysis is *alternatives evaluation*, largely conducted after the alternatives assessment is completed. It is a value-based balancing of the respective attributes (e.g., lower toxicity vs. higher cost) of the regulated product and of the alternatives. Its goal is to essentially rank the regulated product and alternatives in the relative order of how well each option fits the decision criteria guiding the evaluator.

Figure I-1
Alternatives Analysis Framework

Alternatives Assessment

- Develop generic alternatives assessment matrix including higher level criteria (e.g., health, safety, environmental impact, technical performance, economic feasibility), sub-criteria, and specific measurement criteria.
- For regulated hazardous product or process, identify potential alternatives.
- Collect data on regulated product and alternatives
- Transform data to compare attributes

→

Alternatives Evaluation

- Identify evaluation methodology for comparing regulated products and alternatives
- Develop and apply weightings for all level of criteria
- Apply evaluation methodology to compare regulated product and alternatives

Alternatives analysis can be extremely challenging even in relatively straightforward circumstances. The statute requires consideration of a broad range of criteria, some of which are qualitative in nature and many of which are incommensurable. Cognitive psychology and decision theory both recognize that humans have substantial difficulty managing and synthesizing such diverse, rich streams of information in a consistent and coherent fashion.[11] We

[10] Peter J. Sinsheimer and Timothy F. Malloy, Integrating *Safer Alternatives into Chemical Policy: Developing a Regulatory Framework for AB 1879* (Sustainable Technology & Policy Program 2009).
[11] *See* Gregory A. Kiker, *et al.*, Application *of Multicriteria Decision Analysis in Environmental Decision Making*, 1 Integrated Environmental Assessment and Management 95 (2005); T.L. McDaniels, et. al., *Democratizing Risk*

adapt by adopting cognitive heuristics—decision-making rules of thumb that tend to simplify data and emphasize biases. In the regulatory setting concerns regarding consistency and transparency are also raised; ill-considered or even self-interested decision-making can flourish in the shadows of a complex decision environment.

Decision analysis tools can assist policy-makers and stakeholders facing such complex decision environments. One type of decision analysis aid that has been used increasingly in many fields is multi-criteria decision analysis (MCDA). MCDA methods allow decision-makers to assess the performance of different alternatives in a clear, formal way. MCDA is not intended to supplant the decision-maker, nor is it a "black box" from which a selected alternative emerges. Rather it presents a comparative evaluation of the alternatives based upon the criteria provided by the decision-maker, taking into account the relative importance of those criteria to the decision-maker. Properly used, MCDA methods will provide consistency across cases, and are transparent to stakeholders. There are numerous approaches that all fall under the umbrella of MCDA, each involving different protocols for eliciting inputs, structures for representing them, algorithms for combining them, and processes for interpreting and using formal results in actual advising or decision making contexts. A review of articles regarding MCDA use in environmental management from 1990 to 2010 concluded that researchers and policymakers have used MCDA techniques in evaluating environmental alternatives in many areas, including natural resource management, remediation of contaminated sites, the optimization of water and coastal resources, and the strategic planning, stakeholder engagement. Many papers reviewed concluded that the use of MCDA methods provides a significant improvement in the decision process and public acceptance of the suggested remedial or abatement policy.[12]

Even though there are many different MCDA methods and tools, most approaches share a common overall structure. That structure includes the following steps, which closely mirror the content of alternatives assessment and alternatives evaluation:

- **Problem Structuring:** Initially, the problem is defined in terms of relevant stakeholders and overall structure, but is not yet described quantitatively. With this conceptual problem model in mind, the problem is fleshed out by defining alternatives (the potential options the decision-maker is considering) and criteria (the set of properties—such as cost or environmental impact—that describe alternative performance). The criteria are what are used to decide among alternatives. The criteria may be broken into a series of specific sub-criteria as needed to capture different aspects of those higher level criteria. For example, in the context of AB 1879, within the human health impacts criterion, a chemical/product might be compared to alternatives with respect to series of sub-criteria, including carcinogenicity, reproductive toxicity, developmental toxicity, and other properties.

Management: Successful Public Involvement in Local Water Management Decisions, 19 Risk Analysis 497 (1999); D. Kahneman and A. Tversky. JUDGEMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES (1982).
[12] Ivy B. Huang, Jeffrey Keisler, and Igor Linkov, *Multi-Criteria Decision Analysis in Environmental Sciences: Ten Years of Applications and Trends,* Science of the Total Environment (2011 in press).

- **Model Building**: Metrics are assigned to the criteria and sub-criteria. The metrics, which may be quantitative or qualitative, are used to score each of the alternatives. For example, an "acute toxicity" sub-criterion may be given a quantitative metric such as the LD-50 level available in published literature, or a qualitative metric such as high, medium or low toxicity. In this step, decision makers (or stakeholders) also *weight* the criteria according to the value or importance they put on each criterion. Together, the metrics and weights provide information regarding how well each alternative performs on each criterion and how much we care about that performance.

- **Model Application**: The inputs—the criteria weights and alternative scoring—are used in an MCDA model to essentially rank the alternatives according to the data given.[13] As described in Section II.D, below, each MCDA technique works in a slightly different way, but generally speaking each combines the weights and scoring information to rank the alternatives in terms of how well they perform overall. In addition, MCDA tools can be used to examine why the alternatives fared as they did, and to explore how changes in scoring or weighting would affect the outcome. In other words, MCDA allows the user to understand the decision-making process better, and to identify the most relevant differences in scoring alternatives or weighting criteria among stakeholders.

- **Planning and Decision-Making**: Once the model has been run, the results can be used to make decisions regarding the relative safety and viability of the alternatives, as well as the appropriate regulatory response.

Within the broader goal of exploring an alternatives analysis methodology, the project has two related specific objectives. First, the project develops an example of a rigorous, workable alternatives assessment component—essentially problem structuring and model building—consistent with the direction provided by AB 1879. This includes selecting the attributes to be compared, developing the metrics to be used in that comparison, and identifying and collecting the relevant data. For some attributes, the metrics are largely quantitative while in others qualitative measures are developed.

Second, the project develops an appropriate, practical alternatives evaluation component (*e.g.*, engages in model application). Alternatives evaluation requires both clear decision rules and the capacity to make trade-offs within general criteria (for example, within the human health criteria comparing carcinogenicity with endocrine disruption) and between them (such as balancing an adverse health impact against an environmental impact.) The balancing of such incommensurables is by nature a subjective process driven by the decision rules and values under which a decision maker is operating. In a regulatory program, those rules and values must be transparent, should reflect AB 1879's language and purposes, and take into account the values of the public, the affected communities and industries, and other relevant stakeholders.

[13]Ranking need not be complete; i.e., a ranking in which all alternatives are ranked relative to one another. For example, some MCDA methods simply identify the "best" alternative, or a set of acceptable alternatives. *See* Jyri Seppala, *et al. Decision Analysis Frameworks for Life-Cycle Impact Assessment*, 5 Journal of Industrial Ecology 45 (2002). The methods used in this project allow for a complete ranking, but can also be used to develop partial rankings or for screening purposes.

Accordingly, in developing an alternatives evaluation approach, the project focuses particularly on two tasks. It examines the impact of weighting the decision criteria used in evaluating the regulated product/process and its alternatives, including methods for assessing and, to the extent appropriate, incorporating stakeholder values in the decision-making process. It also demonstrates the value and limitations of using software-based multi-criteria decision analysis tools to assist in the evaluation. There are a variety of such tools available, reflecting a range of theoretical foundations. The project examines the application of two primary MCDA tools—multiple attribute utility theory (MAUT) and outranking. Using the decision-support software DECERNS,[14] the project applies MAUT and outranking to each of two case studies: solvents used in professional garment cleaning and lead solder used in electronics. The project then evaluates the outcomes under each MCDA approach in each case study under a series of scenarios described in detail below.

The project is thus a feasibility study intended, among other things, to explore the potential use of MCDA in implementing the alternatives analysis requirement of AB1879. As such, and given the relatively short project schedule of nine months, the methodologies used are commensurately limited in scope. For example, as described below, in performing the alternatives assessments for the two case studies we did not generate primary data regarding the baseline chemicals or the alternatives, relying instead upon existing, publically available data. Likewise, in developing weights for the various criteria, we engaged in a streamlined stakeholder elicitation process to explore potential differences between interested groups. Thus, the project is not intended to provide a comprehensive alternatives analysis process, but rather to examine feasibility of particular potential approaches for such a comprehensive process.

II. Methodology

A. Introduction

This section describes the methodologies used in the project. The discussion of methodologies is broken into three segments. The first segment focuses upon the selection of the case studies, explaining the purpose and criteria for selection. The second segment turns to alternatives assessment, including identification of the relevant criteria and the selection of the metrics for those criteria. The last segment addresses alternatives evaluation—balancing of the performance of the regulated chemical/product and the alternatives, respectively, across the criteria to develop a ranking. It involves development of weights for the relevant criteria and the application of MAUT and outranking decision analysis support tools.

One threshold issue bears discussion before engaging in a detailed explanation of the multiple methods used in this project. AB 1879 is silent as to whether DTSC or individual manufacturers (or consortiums of manufacturers) should perform the alternatives assessment. Likewise, the statute provides no guidance regarding whether alternatives analyses should be manufacturer-specific (regardless of who performs them) or instead be sector-based (i.e., a single, centralized alternatives analysis covering a product produced by numerous entities).

[14] *See* http://www.decerns.com for more information regarding DECERNS.

The structural decisions concerning the identity of the responsible entity (private or government) and the scope of the analysis (individual versus sector-based) will have significant implications for the nature of the alternatives analysis process. For example, if individual firms are responsible for evaluating only their own products, it may well be impractical to include broad-based macroeconomic issues such as effects on employment within California as part of the economic impact analysis. In contrast, in the event the state is performing the alternatives analysis for a manufacturing sector, broader socio-economic analysis may be necessary, if not in the alternatives analysis phase then in the evaluation of regulatory responses.[15] Prior draft regulations under AB 1879 and discussions at meetings of the Green Ribbon Science Panel suggest that DTSC is thus far treating the alternatives analysis process akin to an individualized permitting program. Accordingly (and in light of the limited scope of this pilot study), for our purposes we have made the same assumption.

B. Case Study Selection

We reviewed an array of existing alternatives assessments from government, NGOs and academia with a view to selecting two case studies that met specific key criteria:

- The assessment should be one in which one or more potentially feasible alternatives have been identified for a chemical of concern. Alternatives could include drop-in chemical substitutes, material substitutes, changes to manufacturing operations, and changes to component/product design.
- The assessment should focus on a chemical in commerce for which sufficient data were available on health impact, environmental impact, cost, and performance both for the chemical itself and identified alternatives.
- Assessment criteria for which data were available should map to a sizeable portion of the AB 1979 criteria.

Case studies that were considered included several from the US EPA's Design for Environment Program (DfE), the Massachusetts Toxics Use Reduction Institute's (TURI) Five Chemicals Study, and assessments performed by Clean Production Action and the Lowell Institute for Sustainable Production on deca-BDE in electronics and textiles. Alternatives analyses performed by research team members in various consumer product and service sectors were also considered for inclusion. In particular, the review included:

[15] If one assumes a quasi-permitting scenario in which individual businesses perform AAs, then consideration of economic impacts should be limited to the costs/revenues, if any, experienced by the individual firm in adopting an alternative. Thus, the analysis should not consider the broader impacts on the economy in terms of jobs, business growth, etc. This is consistent with economic cost approaches used in permitting programs generally (such as new source review) and in REACH for the authorization process. The analysis should take the form of a comparative assessment of the net present value of the direct and indirect costs/revenues associated with the baseline and alternative products. The analysis should also consider impacts on the price to consumers. Useful examples are set out in the REACH Authorization guidance and EPA's Cleaner Technologies Substitutes Assessment methodology. (*See* ECHA, *supra* n. 8); EPA, CLEANER TECHNOLOGIES SUBSTITUTES ASSESSMENT-- A METHODOLOGY AND RESOURCE GUIDE (EPA/744-R-95-002 Dec. 1996).

- **The US EPA DfE Lead Free Solder Partnership (2002-2005).** EPA noted that this project "used a life-cycle assessment approach to examine the impacts of tin-lead, tin-copper, tin-silver-copper, and tin-silver-copper-bismuth solders".[16] The project evaluated health and environmental impacts of tin/lead solder and selected lead-free alternative solders, including recycling and reclamation at the end of the electronic product life-cycle, and the leachability of lead-free solders and their potential environmental effects.

- **The US EPA DfE Furniture Flame Retardancy Partnership (2003-2005).** The project report for this partnership, "Environmental Profiles of Chemical Flame-Retardant Alternatives for Low-Density Polyurethane Foam "contains a summary of the environmental and human health attributes of selected flame retardants as alternatives to pentabromodiphenyl ether (pentaBDE)."[17]

- **The US EPA DfE Computer Display Partnership**, which evaluated the life cycle impacts of CRTs vs. LCD displays. Its 2002 report "generated data to assist original equipment manufacturers (OEMs) and suppliers in the electronics field in incorporating environmental considerations into their decision-making processes and identify areas for improvement. This project combined both the Life-Cycle Assessment (LCA) and the Cleaner Technologies Substitutes Assessment (CTSA) approaches to analyze the environmental impacts, performance, and cost of both CRT and LCD desktop monitors."[18]

- **The Lowell Center for Sustainable Production's** "Decabromodiphenylether: An Investigation of Non-Halogen Substitutes in Electronic Enclosure and Textile Applications" (April, 2005) and **Clean Production Action's** (CPA) "Green Screen for Safer Chemicals: Evaluating Flame Retardants for TV Enclosures" (March 2007).[19] The Lowell Center's study "identifies the primary sectors that use decaBDE in the US and substitute flame retardants and materials in electronic enclosures and textiles that meet fire protection standards, assess the availability of these substitutes, delineates examples of where these substitutes are currently used in commerce, and assess the costs of these substitutes where data are available." The CPA study "evaluated three flame retardants that currently meet performance criteria for use in the external plastic housing of televisions (TVs)."

[16] US EPA Design for the Environment Program, Lead Free Solder Partnership, http://www.epa.gov/dfe/pubs/projects/solder/index.htm
[17] US EPA Design for the Environment Program, Furniture Flame Retardancy Partnership, http://www.epa.gov/dfe/pubs/projects/flameret/index.htm
[18] US EPA Design for the Environment Program, Computer Display Partnership, http://www.epa.gov/dfe/pubs/projects/computer/index.htm
[19] Lowell Center for Sustainable Production, http://www.sustainableproduction.org/downloads/DecaBDESubstitutesFinal4-15-05.pdf; Clean Production Action, http://www.cleanproduction.org/Greenscreen.php

- **The Massachusetts Toxics Use Reduction Institute** (TURI)'s "Five Chemicals Study, 2006: Perchloroethylene in Garment Cleaning"[20] and analysis by Occidental College's Urban and Environmental Policy Institute's **Pollution Prevention Center** of garment cleaning alternatives (2000 to 2009)[21], the **City of San Francisco's Department of Environment's** "Alternatives Assessment of Garment Cleaning Alternatives (2008-2009)."[22]

Applying the case study selection criteria to the available candidates, the project team selected the following two cases.

Case Study 1: Garment cleaning solvents. This case study was selected for the existence of data from multiple sources, including the TURI Five-Chemicals study as well as additional extensive analysis performed by research team member Dr. Peter Sinsheimer while at Occidental College's Urban and Environmental Policy Institute's Pollution Prevention Center, and by research team member Dr. Ann Blake as a consultant to the Toxics Reduction Program of the City and County of San Francisco's Department of Environment. The case study examines the following processes:

Table II-1
Case Study 1: Garment Care Alternatives

Alternative	**Description**
Perchloroethylene (PCE) Dry Cleaning *(Baseline Regulated Process)*	Chlorine-based solvent technology. Washing and drying in the same drum. Utilizes solvent recovery system which condenses heated and vaporized solvent during dry cycle and distillation cycle.
DF-2000	Petroleum-based solvent technology. Washing and drying in the same drum. Utilizes solvent recovery system which condenses heated and vaporized solvent during dry cycle and distillation cycle.
GreenEarth	Siloxane-based solvent technology. Washing and drying in the same drum. Utilizes solvent recovery system which condenses heated and vaporized solvent during dry cycle and distillation cycle.
Rynex	Glycol ether-based solvent technology. Washing and drying in the same drum. Utilizes solvent recovery system which condenses heated and vaporized solvent during dry cycle and distillation cycle.
nPropyl Bromide	Bromine-based solvent technology. Washing and drying in

[20] Toxics Use Reduction Institute, University of Massachusetts Lowell, Five Chemicals Study, http://www.turi.org/library/turi_publications/five_chemicals_study
[21] Occidental College, Urban and Environmental Policy Institute, Pollution Prevention Center http://departments.oxy.edu/uepi/ppc/projects.htm
[22] City and County of San Francisco, Department of the Environment, Toxics Reduction Program, Garment Cleaning, http://www.sfenvironment.org/our_programs/index.html

Alternative	Description
	the same drum. Utilizes solvent recovery system which condenses heated and vaporized solvent during dry cycle and distillation cycle.
Carbon Dioxide	C02-based solvent technology. Washing and drying in the same drum. Utilizes solvent recovery system achieve through a change in pressure. Distillation uses heat and condensation to clean solvent.
Professional Wet Cleaning	Water-based solvent technology. Typically washing and drying in different machines. Solvent recovery not used.

The data set for this case study included detailed human health impact assessment data as well as comparative cost analyses for each of the alternative cleaning solvent options. The TURI study included five categories of PCE alternatives for dry-cleaning, assessing them the basis of health, environmental, physical chemical properties (e.g., flammability), cost and performance criteria.[23] Occidental College's Garment Care Project has published multiple assessments of wet cleaning versus various solvent garment cleaning options since 2000, including on health factors, energy and water use, efficacy and cost.[24] San Francisco's Department of Environment's Toxics Reduction Program's Commercial division partnered with Cal/EPA's Office of Environmental Health Hazard Assessment in 2008 and 2010 on a comprehensive assessment of health and environmental impacts of garment cleaning solvent alternatives. The summary documents of this project also included regulatory considerations (e.g., local fire department and hazardous waste permitting requirements) and costs (e.g. local air district grants for certain types of equipment.)[25]

Case Study 2: Alternatives to lead solder in electronics. This case study was selected for the richness of the data set around both environmental and human health impacts, functional performance of the various alternatives and the life-cycle approach taken to the analysis. In addition, academic analyses of the impact of this alternatives assessment on the electronics industry were available to provide additional data. US EPA contracted with the University of Tennessee's Center for Clean Products and Clean Technologies to produce the life-cycle assessment; the methodology and data are described in great detail in a well-documented report available on the US EPA DfE website. The case study included the following alternatives:

Table II-2
Case Study 2: Bar Solder Alternatives

Alternative	**Description**
SnPb *(Baseline Regulated Product)*	Solder alloy composed of 63% tin/37% lead.
SAC (Water quenched)	Solder alloy composed of 95.5% tin/3.9% silver/0.6% copper. Water quenching used to cool and harden solder.
SAC (Air cooled)	Solder alloy composed of 95.5% tin/3.9% silver/0.6%

[23] TURI, *supra,* n. 20.
[24] Occidental College, *supra,* n. 21.
[25] San Francisco Department of Environment, *supra,* n. 22.

Alternative	Description
	copper. Air used to cool and harden solder.
SnCu (Water quenched)	Solder alloy composed of 99.2% tin/0.8% copper. Water quenching used to cool and harden solder.
SnCu (Air quenched)	Solder alloy composed of 99.2% tin/0.8% copper. Air used to cool and harden solder.

This case study met the selection criteria in that it included data on life cycle impacts of alternatives to bar solder from resource extraction to energy use and other environmental impacts. In addition, the case study data included some human health exposure data for both occupational contexts and exposures to the general population. Available data included information the functional performance and limited data on economic feasibility on each of the alternatives.

C. Alternatives Assessment: Problem Structuring and Model Building

The purpose of creating a generic alternatives assessment model was to create a uniform method for comparing a regulated hazardous product to one or more alternatives in order to determine the safety and viability of these substitutes. Because MCDA tools are specifically designed to handle complex data sets, the generic alternatives assessment model was constrained by neither the number of criteria and sub-criteria nor the form of the data – categorical, ordinal, continuous, and nominal. This created the freedom to construct a generic model consistent with AB1879 and most relevant to the decision-maker in determining whether a safer, viable alternative exists.

The first task in developing the generic alternatives assessment model was to determine the major criteria and lower levels of sub-criteria, including the "measurement" sub-criteria (the sub-criteria to which the metrics are linked) and metrics against which the regulated product/process and alternatives are scored. The next step involves integrating the criteria, sub-criteria and metrics into a "performance matrix," which is populated with data specific to the case under study.

To start developing a general alternatives assessment model, we created a table of criteria based on the fifteen factors identified in Section 25253(a)(2) of AB 1879. Table II-3 uses the wording of criteria as it appears Section 25253(a)(2) of the statute and assembled these measures as major criteria and sub-criteria.

Table II-3
Top Level Criteria for Alternatives Assessment Model Based on AB1879

Major Criteria	Sub-Criteria
Public health impacts (K)[26]	• Potential hazards posed by alternatives (Section 2)[27] • Critical exposure pathway (Section 2) • Potential impacts to sensitive subpopulations, including infants and children (K)
Environmental impacts (L)	• Potential hazards posed by alternatives (Section 2) • Critical exposure pathway (Section 2) • Materials and resource consumption (C) • Water conservation (D) • Water quality impacts (E) • Air quality (F) • Production, in-use, and transportation energy inputs (G) • Energy efficiency (H) • Greenhouse gas emissions (I) • Waste and end-of-life disposal (J)
Product function or performance (A)	• Useful life (B)
Economic impacts (M)	• Useful life (B)

In assembling Table II-3, it was clear that certain sub-criteria could apply to more than one major criterion. Because hazards and exposure pathways can apply to both adverse public health impacts and environmental impacts these were listed as sub-criteria for both of these major criteria. Similarly, useful life can be considered both as a performance measure as well as an economic measure. Likewise, it was clear that most of the sub-criteria are not parameters that can be directly measured and therefore need further refinement in order to identify specific measures of these criteria. It was also clear that economic impact and product function/performance sub-criteria need to be developed, given that useful life was the only sub-criteria listed that related to either of these two broader criteria.

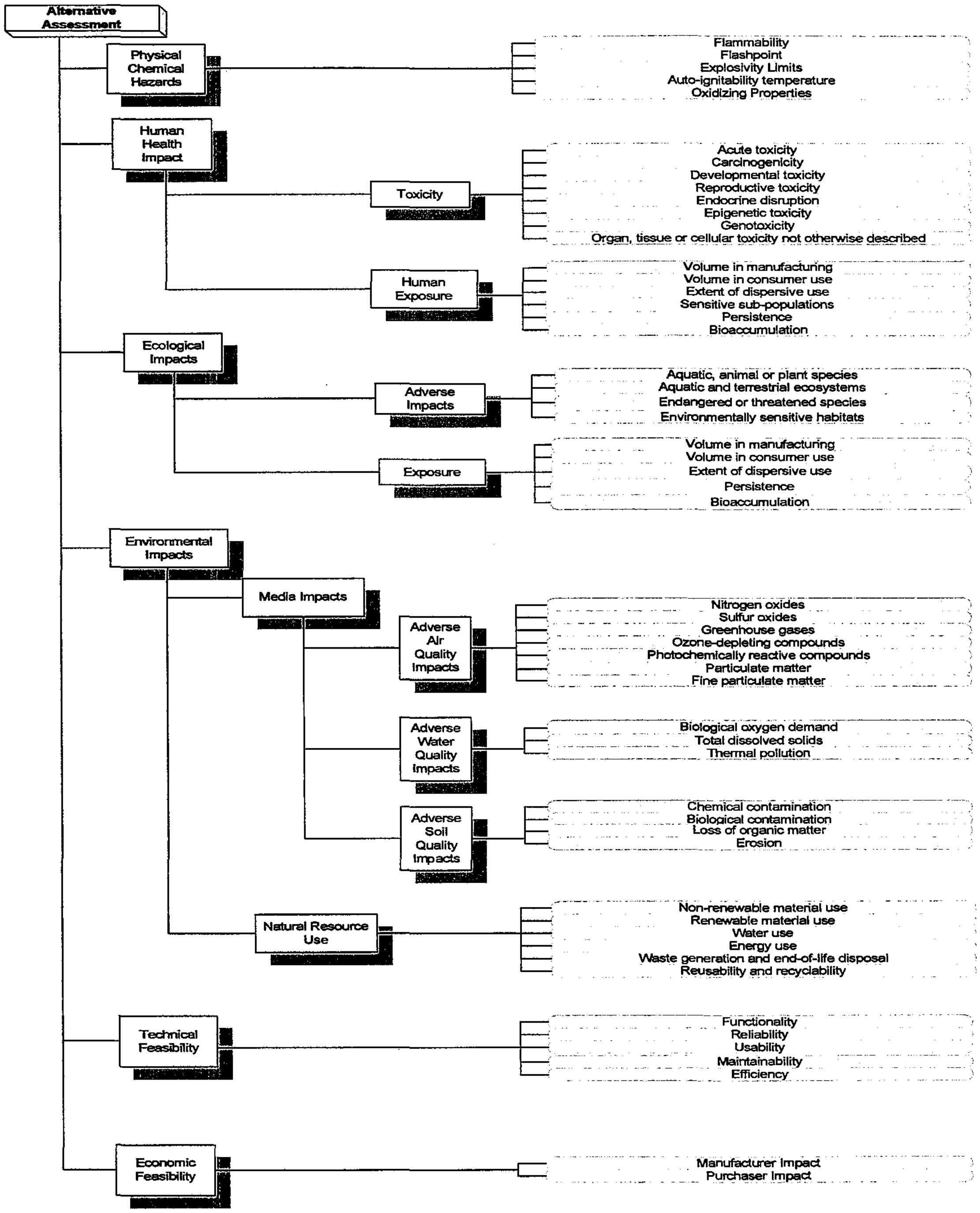
Taking Table II-3 as a starting point, we made a number of modifications and additions in generating a final general alternatives assessment model. The revisions were based upon the evaluation of other alternatives assessment approaches we reviewed, the experience and expertise of project team members in conducting alternatives assessment research, the draft regulations promulgated by DTSC under AB 1879, and discussions within the research team.

Figure II-1 sets out the final generic alternatives assessment model. The generic model consists of seven major comparison criteria. Two new major criteria were added from the original model – physical/chemical hazards and ecological impact. Three of the major criteria

[26] Letters in parentheses indicate the subsection of the Section 25253(a)(2) from which the criteria of sub-criteria were drawn.
[27] "Section 2 refers to Section 2 of AB 1879.

(physical/chemical hazards, technical feasibility, and economic feasibility) lead directly to *measurement sub-criteria,* defined as criteria having metrics against which alternatives can be scored. For example, one measurement sub-criterion associated with physical/chemical hazards is flashpoint, a criterion that can be measured by means of specified tests. Two other major criteria, human health impact and ecological impact, are composed of two intermediate sub-criteria which in turn lead to a set of measurement sub-criteria. One major criterion, environmental impacts, is comprised of two intermediate sub-criteria, with one branch (natural resource use) leading directly to a set of measurement sub-criteria and the other branch (media impacts) leading to three other intermediate sub-criteria representing air, water, and soil impacts, each of which leads to their own set of measurement sub-criteria.

Figure II-1
General Alternatives Assessment Model – Upper Level Criteria and Sub-Criteria



Below is an explanation of each of the major criteria, the sub-criteria branching from it, and a discussion of how this section of the model was developed.

Physical Chemical Hazards

Physical Chemical Hazards criterion was added as a top-level criterion. There was a strong justification for adding this criterion given that fire, explosion and similar hazards are traditionally considered a separate risk from human health hazards or environmental hazards.[28]

Five measurement sub-criteria were identified as physical chemical hazards – flammability, flashpoint, explosivity limits, auto-ignitability temperature, and oxidizing properties. Each of these hazards is associated with fire and explosions, which can result in both human health harm and ecological harm. There was a discussion about whether to make physical chemical hazard a sub-criterion within both human health and ecological impacts. Yet, the human health or ecological harm caused by these physical chemical hazards did not result from direct exposure to the chemical but was caused an indirectly for fire or explosion.

Human Health Impact

In Table II-3, Public Health Impact was identified as a major criterion specified in the statute and three sub-criteria stated in the statute were linked this criteria – hazards, exposure, and impacts on sensitive sub-populations. In the generic model, this criterion was renamed as Human Health Impact. The generic model is links two sub-criteria to human health impact – Toxicity and Human Exposure. Toxicity is linked to eight measurement sub-criteria and Human Exposure is linked to six measurement sub-criteria.

We placed Human Exposure as a sub-criterion of Human Health Impact parallel to sub-criteria of Toxicity. This was based upon traditional notions of risk; namely, that the ultimate impact of chemical depends upon both its inherent hazard and the level and nature of exposure. For example, in comparing two alternatives, a decision-maker may be more concerned about a moderately toxic chemical with high exposures than about a more toxic chemical having extremely low exposure.[29] The generic model captures this notion. That said, this design treats the relationship between hazard and exposure very generally; it does not link a specific exposure concern (*e.g.*, ingestion by an infant) to the hazard criteria with which it is associated (*e.g.*, developmental toxicity). An alternative approach worthy of future consideration would link exposure concerns more directly to the specific hazards, perhaps by developing an integrated metric reflecting both.

In terms of exposure measurement sub-criteria, we used the volume of chemicals in manufacturing, volume of chemicals in consumer use, and extent of dispersive use as surrogate indicators of the likelihood of occupational, bystander and consumer exposure.[30] Persistence and bioaccumulation were selected as human exposure measures because an increase in either

[28] OSHA, Recommended Format for Material Safety Data Sheets (MSDSs), http://www.osha.gov/dsg/hazcom/msdsformat.html.

[29] We acknowledge that this formulation injects the notion of risk into the analysis, and that one could create a model that focuses exclusively on hazard. However, the statute itself incorporates exposure as a factor to be considered.

[30] We noted that the European Union uses these factors as well in prioritizing chemicals as Substances of Very High Concern under REACH. *See* ECHA, General Approach for Prioritisation of Substances of Very High Concern (SVHCs) for Inclusion in the List of Substances Subject to Authorisation (May, 28, 2010).

increases the likelihood and scope of exposure to humans. Exposure of sensitive sub-populations was placed under exposure because the increased concern evident in the statute regarding such exposures.

Ecological Impacts

In the initial alternatives assessment model set out in Table II-1, all environmental parameters listed in AB 1879 were grouped under the criteria "Environmental Impacts." The broad reach of this grouping obscured the distinction between negative impacts to animals, plants and the ecosystems in which they exist on the one hand, and other generalized environmental impacts on the other. For example, impacts to animals incorporates the same notions of hazard and exposure as human health impacts. Accordingly, we split out Ecological Impacts from the broader Environmental Impacts (which are discussed separately below.)

The upper level Ecological Impacts criterion is linked to two sub-criteria, Adverse Impacts and Exposure. Adverse Impacts were linked to four measurement criteria: Aquatic, animal or plant species; Aquatic and terrestrial ecosystems; Endangered or threatened species; and Environmentally Sensitive Habitats. Exposure was linked to five measurement criteria: volume of the chemical in manufacturing, volume in consumer use, extent of dispersive use, persistence, and bioaccumulation.

Environmental Impacts

Environmental Impacts encompassed eight of the fifteen measures specified in AB 1879. These eight measures fell into two distinct sub-criteria: Natural Resource Use and Media Impacts. Media Impacts was further broken down into three additional sub-criteria: Adverse Air Quality Impacts, Adverse Water Quality Impacts, and Adverse Soil Quality Impacts.

Natural Resource Use was linked to six measurement sub-criteria: non-renewable material use, renewable material use, water use, energy use, waste generation and end-of-life disposal, and reuseability and recyclability. Five measures listed in AB1879 were used as a basis for these measures. While one measure (waste generation and end-of-life disposal) was taken directly from the statute, the others were modified to fit the generic model constraints.[31]
For our generic model, Adverse Air Quality Impact was determined by seven measurement criteria. Five of the measures – nitrogen oxides, sulfur oxides, greenhouse gases, ozone-depleting compounds, and particulate matter – were identical to measures listed for this sub-criterion in the June 2010 DTSC proposed regulation. In addition, we added photochemically reactive compounds and fine particulate matter. Measures in the June 2010 proposed regulation

[31] We assumed that the AB 1879 sub-criterion "water conservation" was analogous to the sub-criterion "water use" – technologies using less water could be defined as water conserving. While AB1879 lists two sub-criteria related to energy – "Production, in-use, and transportation energy inputs" and "Energy efficiency," the sub-criterion of "energy use" captures both of these. The remaining AB1879 listed measure in this category is "materials and resource consumption." Because energy and water were already specified, we took this measure to mean the physical materials used during the product's life cycle. The measurement sub-criteria Renewable Material Use and Non-Renewable Material Use cover the product life cycle from extraction, to production, and through the product's lifetime use. The measurement sub-category Reusability and Recyclability covers the product at the end of its useful life.

not included here are toxic air contaminants (which was integrated into human health impacts in this model), secondary organic aerosols, and other ozone forming compounds (generally covered under photochemically-reactive compounds).

Adverse Water Quality Impacts was defined by three measurement sub-criteria: biological oxygen demand, total dissolved solids, and thermal pollution. While all three of these were identified in the June 2010 DTSC proposed regulation, seven additional measures listed in this document were not included in our model. Five of these seven referred to toxicity and therefore would be covered under Ecological Impacts.[32] As for the other two, one was already adequately covered (chemical oxygen demand) and the other was deemed unnecessary (other impacts).

Adverse Soil Quality Impact was measured by four sub-criteria: chemical contamination, biological contamination, loss of organic matter, and erosion. All four of these were identified in the June 2010 DTSC proposed regulation.[33]

Technical Feasibility

The generic alternatives assessment model uses five measurement sub-criteria to define the technical feasibility of alternatives relative to the regulated product: Functionality, Reliability, Usability, Maintainability, and Efficiency. These criteria were defined as follows: Functionality – a set of functions that satisfy the stated or implied need; Reliability – attributes that affect the capability of the product to maintain its level of performance; Usability – attributes that bear on the effort needed to use the product; Maintainability – the ability to identify and fix faults in the product; and Efficiency – attributes that bear on the resources needed to use the product.

These measures were taken from the International Standardization Organization (ISO) 9126 standard for determining the quality of software products.[34] Each of these five criteria appears to be readily generalizable to any consumer product.[35] AB1879 provides little guidance on how to measure technical feasibility, listing only two factors that related to this criterion: Production Function or Performance and Useful Life. The ISO measurement criteria clearly cover the both of these stated parameters. Three of the ISO measures relate to the product's function – functionality, usability, and efficiency while two relate to performance – reliability and maintainability. Useful Life can readily be conceived as an attribute of reliability.

---

[32] These five include: Chronic and acute toxicity in the water column and sediments; Chemicals identified as priority toxic pollutants for California pursuant to section 9 303(c) of the federal Clean Water Act and listed in section 131.38 of Title 40 of the Code of Federal Regulations published in the Federal Register May 18, 2000; Pollutants listed by California or the United States Environmental Protection Agency for one or more water bodies in California pursuant to section 303 (d) of the federal Clean Water Act; Chemicals identified as contaminants that have primary Maximum Contaminant Levels (MCLs) under the federal Safe Drinking Water Act; Pollutants requiring monitoring and reporting in waste discharges to land that have Notification Levels (NLs) specified under the Waste Discharge and Water Reuse Requirements (WDRs/WRRs) of the Porter-Cologne Water Quality Control Act.

[33] Loss of biodiversity; Compaction or other structural changes; Soil sealing; Other impacts the affect or alter soil function or soil chemical, physical or biological characteristics or properties.

[34] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=16723.

[35] Portability, a sixth measurement criterion, was not as readily generalizable and therefore was not included.

Economic Feasibility

Economic feasibility in the generic alternatives assessment model was defined by two measurement criteria: Manufacturer Impact and Purchaser Impact. Manufacturer Impact refers to the extent to which expected revenues in selling a product are greater than expected costs of manufacturing the product, taking into account the manufacturer's ability to pass on increased costs to the consumer or to its suppliers. Purchaser Impact refers to the increased price paid by the consumer for the end product.

As with technical feasibility, AB1879 provides little guidance on how to measure economic feasibility, listing only two factors that could directly relate to this criterion: Economic Impacts and Useful Life. The European Chemicals Agency (ECHA) developed a well-conceived approach to economic feasibility in its guidance document for authorization of chemicals within the REACH regulation.[36] One measurement criterion for economic feasibility was "whether the net present value of the revenues minus costs is positive" taking into account whether it is possible to pass-through costs to consumers. The project incorporates these notions in its sub-criteria for economic feasibility. When determining the economic feasibility of alternatives in the context of authorization, ECHA recommends that the measurement criteria focus on product-specific factors related to costs and revenues and not broader macroeconomic factors.

D. Alternatives Evaluation

As discussed above, the alternatives evaluation component of alternatives analysis (analogous to "model application" in decision theory) includes the development of weights for the decision criteria, and the evaluation of the relative performance of the regulated product and alternatives regarding those criteria. This section describes the methods the project used in performing those two steps, beginning with weighting.

D.1 Weighting and Stakeholder Elicitation

In most situations decision-makers are not equally concerned about all decision criteria. For example, an individual decision-maker may place more importance on whether a household cleaner causes cancer than on whether it contributes to smog formation. Accordingly, the decision-making method should take the respective importance or weight placed upon the criteria into account in evaluating alternatives. Given the fact that individuals may place different weights upon criteria, the concept of weighting raises significant questions in the context of a regulatory program. For example, to what extent should the program take into account the potentially inconsistent weighting preferences of the various stakeholder groups and of the public at large? Likewise, what role should pragmatic and strategic considerations of the regulators play in the weighting?[37]

---

[36] ECHA, *supra*, n. 8.

[37] *See* United Kingdom Department for Communities and Local Government, MULTI-CRITERIA ANALYSIS: A MANUAL (January 2009).

Existing regulatory programs that involve multi-criteria decision-making, such as Superfund[38], SNAP[39] and CEQA[40], do not explicitly confront the question of weighting. Instead, weighting is typically performed on a largely *ad hoc* basis, generally without any direct, systematic discussion of the relative weights to be accorded the relevant decision criteria. Such *ad hoc* treatment of weighting raises serious concerns regarding the consistency of outcomes across different cases. Over time regulators may develop standard outcomes or rules of thumb which provide some consistency in outcome, but such conventions and the tacit weighting embedded in them undermine transparency in decision-making. Moreover, without clear guidance regarding the relative weight to be accorded to criteria, government decision-makers may consciously or inadvertently allow arbitrary political or administrative factors to influence the decision. Finally, without a transparent process of determining the viability of alternatives, industry is not able to predict how an agency might rule making it difficult to plan manufacturing an existing product line or funding research and development on safer substitutes.

Generally speaking, criteria weights can be established in one of three ways: use of existing generic weights, calculation of weights using objective criteria, or elicitation of weights from experts or stakeholders.[41] Generic weighting can be as simple as applying equal weights to all criteria, or relying upon publically available sets of criteria weights developed by third parties through calculation or elicitation. For example, the National Institute of Standards and Technology (NIST) provides four generic weighting sets as part of its BEES life cycle impact assessment software for the selection of environmentally preferred building products, set out in Table II-4. [42] The criteria in Table II-4 reflect a set typically used in life cycle impact assessment.

---

[38] This is the federal hazardous site cleanup program created under the Comprehensive Environmental Response, Compensation, and Liability Act of 1980, 42 U.S.C Section 9601, *et seq.* In that program, EPA chooses cleanup options using nine decision criteria.
[39] *See* Section I.A, above.
[40] California Environmental Quality Act, California Public Resources Code Section 21000 *et seq.*
[41] USEPA, FRAMEWORK FOR RESPONSIBLE ENVIRONMENTAL DECISION-MAKING (FRED): USING LIFE CYCLE ASSESSMENT TO EVALUATE PREFERABILITY OF PRODUCTS (EPA/600/R-00/095 October 2000); Xiaoying Zhou and Julie M. Schoenung, An Integrated Impact Assessment and Weighting Methodology: Evaluation of the Environmental Consequences of Computer Display Technology Substitution, 83 J. Envt'l Man. 1 (2007). For a critique of weighting methods in life cycle impact assessment generally, *see* Wulf-Peter Schmidt and John Sullivan, *Weighting in Life Cycle Assessment in a Global Context*, 7 International Journal of Life Cycle Assessment 5 (2002).
[42] Thomas P. Gloria, Barbara C. Lippiatt, and Jennifer Cooper, *Life Cycle Impact Assessment Weights to Support Environmentally Preferable Purchasing in the United States*, 41 Environ. Sci. Technol. 7551 (2007).

Table II-4
NIST Optional Weight Sets

Criteria	NIST Panel	EPA SAB	Harvard Study	Equal Weights
Global Warming	29.3	16	11	8.3
Fossil Fuel Depletion	9.7	5	7	8.3
Criteria Air Pollutants	8.9	6	10	8.3
Water Intake	7.8	3	9	8.3
Human Health Cancerous	7.6	11	6	8.3
Human Health Non-cancerous	5.3	11	6	8.3
Ecological Toxicity	7.5	11	6	8.3
Eutrophication	6.2	5	9	8.3
Habitat Alteration	6.1	16	6	8.3
Smog	3.5	6	9	8.3
Indoor Air Quality	3.3	11	7	8.3
Acidification	3.0	5	9	8.3
Ozone Depletion	2.1	5	11	8.3

The NIST Panel weights were generated through a facilitated stakeholder elicitation processes involving seven building products manufacturers, seven users, and five life cycle assessment experts.[43] The EPA SAB weights and Harvard Study weights were derived by NIST from sets of qualitative rankings developed by the EPA's Science Advisory Board in 1990 and a team of Harvard researchers in 1992, respectively.[44] We chose not to use existing generic weights for several reasons. First, neither the NIST impact categories nor those of other existing weighting schemes map well onto the criteria suggested by AB 1879. Second, the composition of the panels used by NIST does not reflect the diversity of opinion we were seeking. The SAB and Harvard panels were expert-heavy, providing little input from industry, government or NGOs. The panel convened by NIST was broader, including manufacturers and end-users, but nonetheless lacked NGO or regulatory perspectives. Third, the panels had no obvious connection to California and thus were unlikely to reflect the range of social and other values unique to California.

Weights can also be calculated using objective measures. For example, in distance-to-target methods, all criteria are initially assumed to be of equal importance, and then each is weighted by reference to the magnitude of variance between the desired conditions and existing conditions for each criterion. Thus, for example, if one assumes that the global community is further away

[43] *Id.* The facilitator used an analytical hierarchy process (AHP).

[44] Barbara Lippiat, The BEES Model for Selecting Environmentally and Economically Balanced Building Products (1997) http://www.fire.nist.gov/bfrlpubs/build98/PDF/b98032.pdf (last visited May 24, 2011); *see* Gloria, *et. al, supra* n. 42; Vicki Norberg-Bohm, *et. al*, INTERNATIONAL COMPARISONS OF ENVIRONMENTAL HAZARDS: DEVELOPMENT AND EVALUATION OF A METHOD FOR LINKING ENVIRONMENTAL DATA WITH STRATEGIC DEBATE MANAGEMENT PRIORITIES FOR RISK MANAGEMENT, Center for Science &International Affairs, John F. Kennedy School of Government, Harvard University (1992)(Harvard Study); U.S. Environmental Protection Agency Science Advisory Board, REDUCING RISK: SETTING PRIORITIES AND STRATEGIES FOR ENVIRONMENTAL PROTECTION (SAB-EC-90-021 1990)(EPA SAB).

from achieving its goal for global warming than it is for ozone depletion, greater weight would be given to global warming potential.[45] Distance –to-target methods have been subject to significant criticism when used alone, primarily because of the underlying assumption that, but for the level of variance from the desired condition, all criteria were of equal concern.[46] Another calculation approach is monetary evaluation, in which weighting is made on the basis of costs related to environmental consequences. While the specific valuations methods vary, they generally rely upon some objective calculation of the costs of responding to or avoiding the criterion's impact. The more significant the monetary value of an assessment criterion, the greater significance it will take on.[47] We rejected the use of these two objective methods because of their limited focus to variance from target and costs, respectively, and their complexity given the limited scope of this project.

The third major approach, the elicitation of weights from experts or stakeholders directly, includes a wide variety of methods, such as public opinion surveys, facilitated group consensus-based procedures (e.g., the modified Delphi technique), and various weighting procedures used in multi-criteria decision analysis models. Given the limited scope of this pilot project, we did not use the resource intensive survey or facilitated consensus-based procedures. MCDA approaches typically use one of three weighting methods: direct rating,[48] pair-wise comparison (an analytical hierarchy procedure often used with outranking methods such as PROMETHEE), and "swing weighting" (typically used with MAUT methods.[49]) We chose direct rating because we needed a simple and transparent weighting approach that could be used for both outranking and MAUT methods. The direct rating approach we used—Max100—is relatively straightforward to apply, and an empirical evaluation of the method demonstrated it to be reliable and preferred by interview subjects.[50]

In light of the narrow scope of this pilot study, the stakeholder elicitation was not intended to develop weightings that were statistically representative of the respective stakeholder groups. Rather, it was designed as an initial exploration of differences among various stakeholder groups regarding weighting, and of the impact of any such differences on the rank order of alternatives. Four stakeholder groups were considered: Environmental Non-governmental Organizations, Industry, Policymakers and Consumers. The elicitation process was also designed to obtain stakeholder reactions to the criteria; for example, whether any relevant criteria have been left out.

---

[45] Zhou and Schoenung, *supra* n. 41, at 17. For an overview of the limitations of distance-to-target methods, see Sebastiao R. Soares, Laurence Toffoletto, and Louise Deschenes, *Development of Weighting Factors in the Context of LCIA*, 14 Journal of Cleaner Production 649 (2006).

[46] Göran Finnveden, *Recent Developments in Life Cycle Assessment*, 91 Journal of Environmental Management 1 (2009).

[47] Soares, *et al.*, *supra* n. 45; Göran Finnveden, A Critical Review of Operational Valuation/Weighting Methods for Life Cycle Assessment (June 1999) .(http://naturvardsverket.se/Documents/publikationer/afr-r-253-se.pdf last visited May 25, 2011). All such methods are not strictly objective; for example, under some approaches, stakeholder elicitation or more general surveys are used to gather information regarding respondents' willingness-to-pay for certain outcomes. *Id.*

[48] Igor Linkov and Emily Moberg, Multi-Criteria Decision Analysis: Environmental Applications and Case Studies (in press 2011); Paul A. Bottomley and John R. Doyle, *A Comparison of Three Weight Elicitation Methods: Good, Better, and Best*, 29 Omega 553 (2001)(describing and comparing direct rating and its variants, Max100 and Min 10).

[49] Valerie Belton and Theodor J. Stewart, MULTIPLE CRITERIA DECISION ANALYSIS (2002).

[50] Bottomley and Doyle, *supra* n. 48.

We identified interview subjects by reviewing the rulemaking docket for the AB 1879 regulations, and considering participation in the Green Ribbon Science Panel meetings. We selected three subjects from each of the four stakeholder groups, and conducted an individual interview with each subject lasting approximately one and a half hours.

The interviewer provided the subject a copy of the conceptual model showing the criteria and sub-criteria, and implemented the following procedure for each level of criteria and sub-criteria in sequence. For each level, the subject received a set of cards (in random order), with each card containing the name of a criteria and a brief definition and example.[51] The subject then organized the cards of criteria by placing them in order, with the most important criteria on the top to the least important criteria on the bottom. If the subject felt that specific criteria should have equal ranking, they paper-clipped those cards together. If the subject felt a criterion was missing, they were asked to add a card for that criterion. The interviewer also asked the subject what factors led the subject to place the criteria in the order they did.

Next the subject indicated the relative importance of the criteria by rating them along a 100 point scale. The interviewer presented the subject with a 100 point scale, noting that the subject's most important criterion (i.e., the first card) would be placed at the 100 mark. Beginning with the subject's most important criterion, the subject then placed each of the criteria along the scale. The subject repeated the card sorting and scaling procedures for each level of sub-criteria until all levels were completed. Time was also allotted for general discussion about the process for Multi-Criteria Decision Analysis, to allow subjects to share their general thoughts, comments and concerns about using this type of a process for implementation under AB 1879.

For each group we used the criteria weights obtained from the interviews to develop the set of average criteria weights using the following equation:

$$w_{i,average} = \frac{\sum_{j=1}^{m} w_{i,j}}{m}$$

Where *m* is the number of interviewees and *i* is the criterion index.

The DECERNS software used in the analysis requires normalized weights as inputs. Normalization was done using the following equation:

$$w_{i,norm\_by100} = \frac{w_{i,average}}{MAX(W_{average})} \cdot 100$$

D.2 Evaluation using MCDA as a Decision-Aid Tool

This project focuses upon two commonly used MCDA methods as tools to assist in the evaluation: multi-attribute utility theory (MAUT) and outranking. We used the software package DECERNS for both MAUT and for outranking, With respect to outranking, DECERNS uses the

[51] The definitions were derived from multiple sources, including EPA, Design for the Environment Program Alternatives Assessment Criteria for Hazard Evaluation (Draft Jan. 2011); OEHHA's draft Hazard Traits regulation, and the Cradle to Cradle Certification Program Version 2.1.1.

outranking method known as PROMETHEE (standing for **P**reference **R**anking **O**rganization **METH**od for **E**nrichment **E**valuations).[52] The discussion that follows provides more detail regarding these two methods. (While this project compares the application of these two methods, selection of one recommended MCDA method in the context of regulatory alternatives analysis is beyond the scope of this project.)[53]

MAUT is an optimization approach, meaning that it represents the decision-maker's preferences as utility functions, and attempts to maximize the decision-maker's overall utility. MAUT is premised on the assumption that the decision-maker has a fairly well-defined set of preferences that can be represented on a dimensionless utility scale. It also assumes that the decision-maker is rational; that is, they prefer more utility rather than less and are consistent in those preferences. In the context of this project, therefore, we generated a utility function for each criterion, which reflects how a decision maker's preference changes for different values of that criterion. This utility function spans from 0 to 1, with a utility of 1 being assigned to the value of the best (or highest) alternative score for that criterion and 0 being assigned to the value of the worst (or lowest) alternative score. In this case, a linear utility function was used, which assumes that increases in utility are directly related to increases in the alternative's score for the criterion in question. We used the linear utility function as a default; in some scenarios below we explore the use of different utility functions that reflect more realistic preferences under some circumstances.

Having converted an alternative's performance on a particular criterion to a score on a 0 to 1 utility scale, we determine the weighted score for that criterion by multiplying the score by the weight assigned to that criteria. Thus, for example, if an alternative performed better than all others on acute toxicity, it would receive a score of 1. If acute toxicity had an overall weight of 0.2, the alternatives score on that criterion would be 0.2. The total score for the alternative is simply the sum of all weighted scores received for all criteria by the alternative. Accordingly, if an alternative preformed the best for *all* criteria, it would receive an aggregate score of 1. Because the weighted scores for all criteria are added to produce the alternative's total score, MAUT is a "compensatory" method. This means that poor performance on one criterion can be compensated by better performance on another.

Outranking models do not create utility functions, but instead directly compare the performance of two alternatives at a time, in terms of each criterion, to identify the extent to which one alternative out-performs the other. It then aggregates that information for all possible pairings to rank the alternatives based on overall performance on all criteria. Generally speaking, the PROMETHEE method used in the project creates a "preference index" for each alternative, which is calculated by reference to the alternative's positive flow (i.e., those instances in which the alternative outperforms another alternative on a given criterion) and negative flow (i.e., those instances in which the alternative is outperformed by another alternative). The value awarded for winning a particular pairing is weighted, meaning it is adjusted to reflect the value placed

[52] *See* Linkov and Moberg, *supra* n. 48.

[53] Commentators have identified a number of factors relevant to such a decision. *See* Thomas P. Seager, Understanding Industrial Ecology and the Multiple Dimensions of Sustainability, in STRATEGIC ENVIRONMENTAL MANAGEMENT 17 (ed. Robert Bellandi 2004); Adel Guitouni and Jean-Marc Martel, Tentative Guidelines to Help Choosing an Appropriate MCDA Method, 109 *European Journal of Operational Research* 501 (1998).

upon that criterion by the decision-maker. Thus, outperforming another alternative in a minor criterion is worth less than outperforming it with respect to a more highly weighted criterion.

As a default in PROMETHEE and most other outranking methods, any difference in performance—however small—will result in an increase in positive flow for the better performing alternative. As in MAUT, PROMETHEE recognizes that a decision-maker may be indifferent to how alternatives perform on certain criteria until certain levels are met or after certain levels are exceeded. For example, taking the example again of acute toxicity, regulators may be indifferent to differences in LD-50 levels above a certain level. Likewise, with respect to economic impact—particularly for materials such as solder which make up a small portion of total production costs—decision-makers may be indifferent to cost impacts until they exceed a threshold. PROMETHEE allows the decision-maker to incorporate such considerations into the generation of the preference index through a variety of preference functions.[54]

Because outranking techniques aggregate the results of pairings for all criteria, they allow superior performance on some criteria to compensate for inferior performance on other criteria. However, they do not necessarily reflect the magnitude of relative underperformance in a criterion versus the magnitude of over-performance in another criterion. In other words, if Alternative A is marginally worse that Alternative B in one criterion, but substantially better with respect to another, outranking may not fully "compensate" Alternative A for its overall better performance. Therefore, outranking models are known as "partially compensatory."

III. Results and Discussion

A. Alternatives Assessment

The baseline performance matrices for the garment care and the bar solder case studies are set out in Appendices A and B, respectively. Each matrix sets out the metrics used for each criteria, as well as the relevant sources and notes. Matrices reflecting other data scenarios discussed below are available from the authors.

The primary data source for lead bar solder case study was an August 2005 EPA document entitled "Solders in Electronics: A Life-Cycle Assessment" which used Life-Cycle Impact Assessment (LCIA) methodology to generate point estimates for a number of human health, ecological, and environmental impacts comparing tin/lead solder with a range of non-lead alternatives.[55] For human health impacts, the LCIA method used in this study focused on two population groups – occupational and public – as well as two impact categories – carcinogenicity and chronic non-cancer effects. Chronic non-cancer effects included reproductive toxicity, developmental toxicity, neurotoxicity, immunotoxicity, behavioral effects sensitization, radiation effect, and chronic effects to other specific organs or body systems.[56] The study did not provide generate impact scores for each individual non-cancer impact, providing only a single score for

[54] Belton and Stewart, supra n. 49.
[55] EPA, Solders in Electronics: A Life-Cycle Assessment, EPA 744-R-05-001, August 2005.
[56] *Id.* at 3-114.

the group as a whole... In addition, occupational scores were based on all hazardous chemicals used throughout the respective product life-cycles of the lead solder and its alternatives.[57]

The EPA study on solder in electronics also provided data on technical performance and economic impact, but this data was extremely limited, and additional sources were identified to fill in data gaps. In addition, while the human health, ecological, and environmental effects were compiled over the life-cycle of the product, performance and economic impacts were limited to the production of the printed circuit board.

Data used to compile the performance matrix for the garment care case study, by contrast, come from a relatively wide variety of governmental, academic, and industry sources. Government sources included the EPA, California's Office of Environmental Health Hazard Assessment, the California Air Resources Board, and the San Francisco Department of the Environment. Academic research came from Occidental College and University of Massachusetts, Lowell. The sources on industry data were from Material Safety Data Sheets provided by chemical manufactures or solvent distributors.

Due to limitations in existing sources, data for this case study regarding human health, ecological, and environmental impacts focused primarily on the solvent as used in cleaning process and not on full life-cycle impacts. Because the garment care case study focused on individual solvents for each technology option, we were able to compile data on a wide range of human health hazards, ecological hazards, and measures of exposure.

For both the garment care case study and the lead solder study, there were missing data (i.e., measures for which there were not data for any alternative) and incomplete data (i.e., measures in which there was some data for a measure but not complete data.) The garment care case study had significantly less missing data (for ten measurement sub-criteria) than the lead solder case study (for thirty-five measurement sub-criteria). Conversely, the lead solder case had no incomplete data, while the garment care study had incomplete data for 17 of the 73 measurement sub-criteria.

For the baseline performance matrices, midpoint values were used fill in both incomplete and missing data. Where there is an external scale used for the criteria metrics (such as the Globally Harmonized System (GHS)), we used the midpoint of that scale. Where there is not external scale, we used the midpoint of the values for all alternatives for which data was available. For weighting, we used the average of all stakeholders interviewed.

B. Alternatives Evaluation

This section presents the results of the alternatives evaluation, beginning with presentation of the weighting regimes derived from the stakeholder elicitation. Next, it describes the results from the evaluation of the "baseline" performance matrices for each case study. It then examines the outcomes from a series of variations from that baseline; namely, variations of certain data assumptions, of weighting, and of the decision-making model itself.

[57] *Id.* at.3-115.

B.1 Weighting and Stakeholder Elicitation

The results of the stakeholder elicitation are set out in Appendix C. The average weights derived for the first level criteria for each of the four groups and for all interviewees are set out in Table III-1 below.[58] As Table III-1 demonstrates, at least at this level, there were not substantial differences across the groups.[59] On average, all stakeholder groups (but Industry) placed more weight on human health and ecological hazards as well as on environmental impact criteria. Industry and Policymakers assigned more weight to technical feasibility as compared to Consumers and Environmental NGOs. Industry placed more weight on economic feasibility than the other three groups. As discussed above, however, the sample sizes for the stakeholder groups were quite small (three in each group), with the goal of getting a sense of the potential differences across and within groups.

Table III-1
Average Stakeholder Weighting

	Envtl. NGO	Industry	Consumer	Policymaker	Overall Average
Physical Chemical Hazards	15.22%	11.04%	15.21%	13.12%	13.75%
Human Health Impact	21.14%	18.07%	20.28%	24.75%	20.83%
Ecological Hazards	18.60%	18.67%	19.68%	18.07%	18.75%
Environmental Impacts	18.60%	20.08%	19.68%	14.11%	18.33%
Technical Feasibility	14.38%	16.47%	11.56%	16.58%	14.58%
Economic Feasibility	12.05%	15.66%	13.59%	13.37%	13.75%

Except as specifically identified below, the evaluation uses the overall average weights derived from the stakeholder solicitation.

B.2 Baseline Scenarios

One primary goal of the project was to demonstrate and examine the operation of two MCDA techniques: MAUT and outranking. The baseline scenarios compare the performance of MAUT and outranking in the context of the baseline performance matrix for each case study. As Figures III-1 and III-2, demonstrate, for garment care, the two MCDA approaches ranked the alternatives in the same order. In both cases, wet cleaning was the best overall performer, followed by CO_2 cleaning and perchloroethylene in that order. Figure III-1 displays the total score received by each garment care alternative under MAUT; the higher the score, the better the overall performance.

[58] The stakeholder elicitation also developed weighting for the lower level sub-criteria. The complete set of weights for all criteria is available from the authors.

[59] While the individual stakeholder elicitation results set out in Appendix C show a fairly wide variation within some of the groups, the ranges of weights for each group do not vary appreciably.

Figure III-1
Garment Care Baseline Outcome Under MAUT

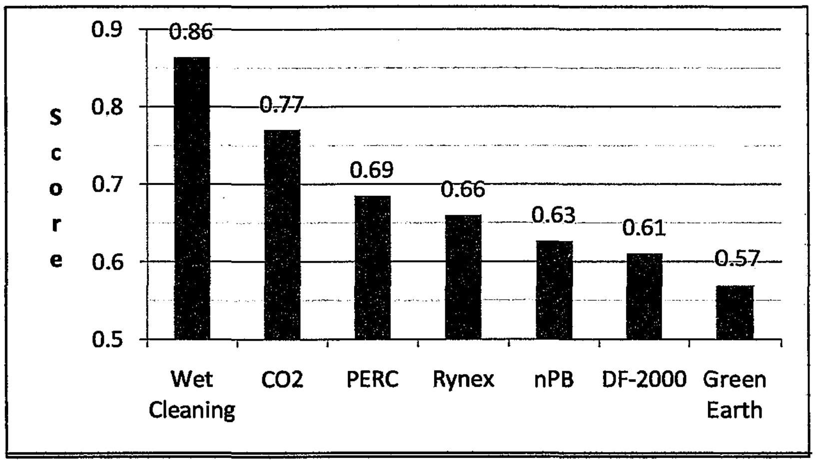
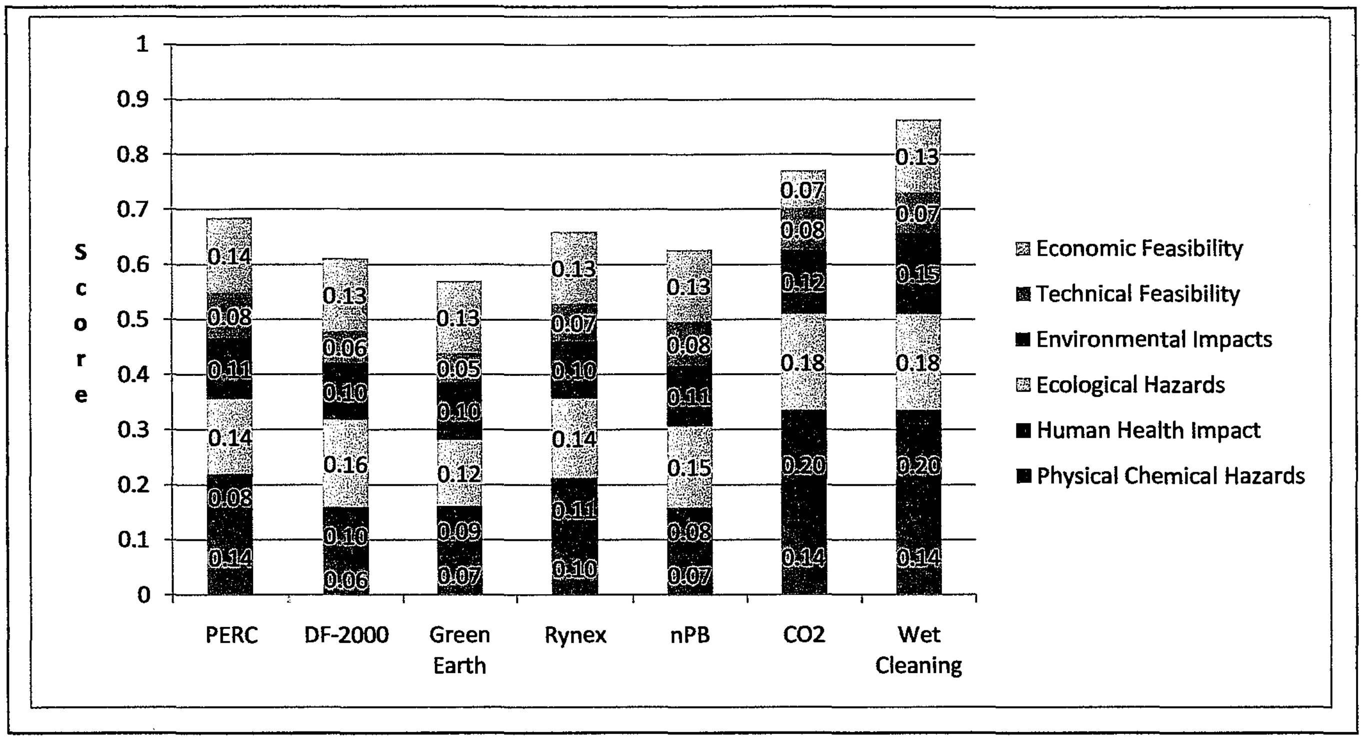


Figure III-2 indicates the relative ordering of the alternatives under PROMETHEE, beginning with the most preferred alternative on the left.

Figure III-2
Garment Care Baseline Outcome Under PROMETHEE

The value of MCDA runs beyond simply generating an ordering of alternatives; the methods also enable decision-makers to understand the basis of the ordering. For example, Figure III-3 breaks down the MAUT score for each alternative to indicate the relative contribution of each criterion to the overall score of the alternative.

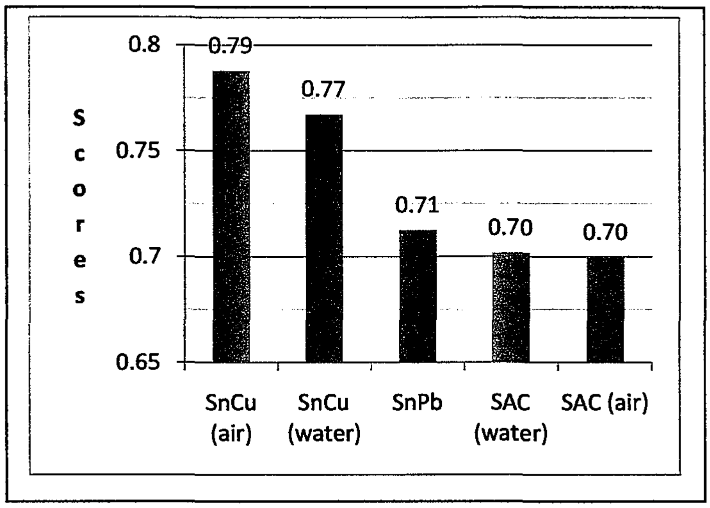
Figure III-3
Garment Care Criteria Contribution to MAUT Scores



This figure demonstrates that, taking into account weighting, wet cleaning and CO_2 cleaning's performance on human health, environmental and ecological criteria drove the outcome in this case. This was so despite CO_2 cleaning's very poor performance in terms of economic impact. Poor performance by DF-2000, nPB, Rynex and Green Earth in terms of physical/chemical hazards placed those alternatives behind the existing technology—perchloroethylene dry cleaning.

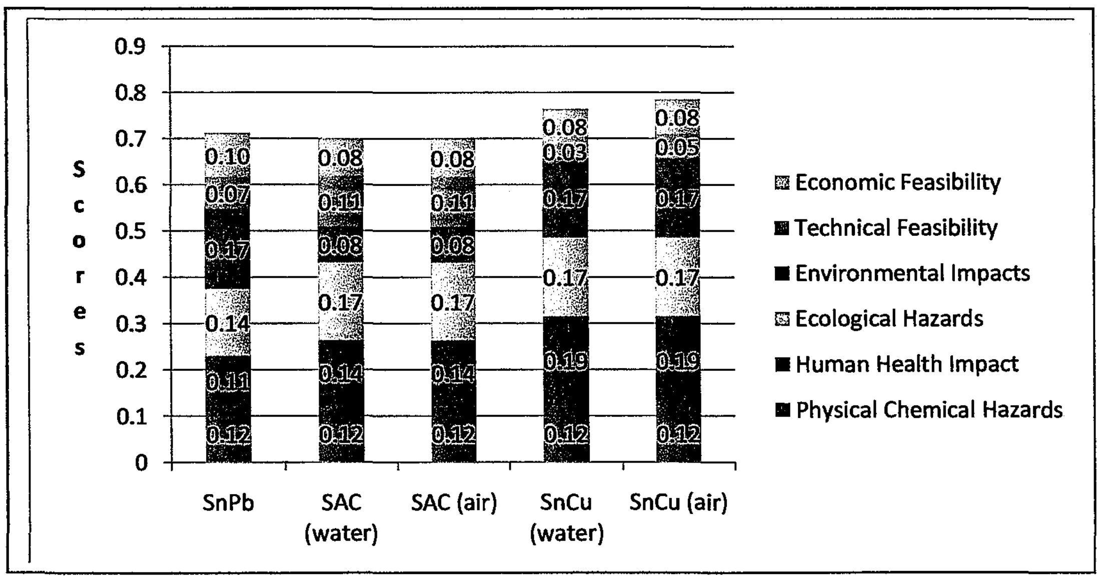
Unlike the case of dry cleaning, complete agreement between MAUT and outranking in the lead solder case was lacking. As Figure III-4 indicates, under MAUT, the two SnCu solders had the highest scores, followed by tin/lead solder, SAC (water cooled) and SAC (air cooled) in that order. However, the tin/lead and SAC solders were quite close in scores.

Figure III-4
Lead Solder Baseline Outcome Under MAUT



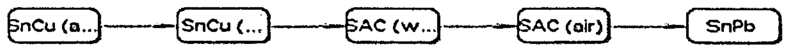
For each alternative, Figure III-5 illustrates the relative contribution of the alternative's performance on each criterion to its overall score.

Figure III-5
Lead Solder Criteria Contribution to MAUT Scores



The order was somewhat different under outranking. As in MAUT, the two forms of SnCu solder were the best performers in outranking. However, unlike MAUT, in outranking tin/lead solder took the last position behind both forms of SAC solder. See Figure III-6.

Figure III-6
Lead Solder Baseline Outcome Under PROMETHEE



As in MAUT, however, the difference between tin/lead, SAC (air cooled) and SAC (water cooled) were relatively small, and there was a noticeably larger gap in performance between this group of three on the one hand and the top two performers on the other.

B.3 Effect of Data Assumption Variations

These baseline results were based on a number of assumptions about the data, including how to deal with incomplete data and missing data and whether keep data as continuous or convert it to categorical.

Each of these assumptions about the data, taken separately or in combination, can be tested within MCDA to see if a different assumption would make a difference with respect to the risk order of alternatives. With this in mind, we developed a set of data scenarios in the garment care case study, summarized in Table III-2. Scenario 1 is the baseline discussed above and is used for comparison with the other scenarios. Scenario 2 is the same as Scenario 1 but removes all criteria in which all data points were missing. In Scenario 3, rather than assume mid-point averages for incomplete or missing data, worst case estimates were assumed for each empty cell. In Scenario 4, all data measured on a continuous scale (such specific temperatures on the Fahrenheit scale) is converted into categorical data (think of "hot," "warm," and "cold").

Table III-2
Baseline and Variations

Scenario Number	Scenario Name	Scenario Description
1	Baseline (Garment Care and Lead Solder)	For incomplete data and missing data, we used a midpoint value. Where there is an external scale used for the criteria metrics (such as the Globally Harmonized System (GHS)), we used the midpoint of that scale. Where there is no external scale, we used the midpoint of the values for all alternatives for which data was available. We assumed a linear relationship between an increase in a score on a criterion and increase score from that criterion.
2	Criteria Removal--Broad (Garment Care and Lead Solder)	Baseline, but for incomplete data, we used a midpoint value. For missing data we removed the sub-criteria from consideration.

Scenario Number	Scenario Name	Scenario Description
3	Worst Case Supplement (Garment Care Only)	For incomplete data and missing data, we used a worst case value. Where there is an external scale, we used the highest negative value of that scale. Where there is no external scale, we used the highest values for any alternative with data available.
4	Categorical Baseline (Garment Care Only)	Baseline but used categorical values rather than quantitative values for flashpoint, acute toxicity, persistence, bioaccumulation, and aquatic toxicity.

Scenario 2: Missing Data. The above baseline results all used a mid-point value for missing and incomplete data. Scenario 2 removes all criteria for which there is missing data (i.e., no data for any of the alternatives). That scenario continues to use the midpoint for incomplete data (i.e., data that is missing for some, but not all, alternatives regarding a single criterion).

For the garment care, removal of criteria for which data was missing had no impact on ranking, and very little impact on alternative scores under MAUT. (See Figure III-7.) For the lead solder case, with three and a half times more criteria missing data, the removal of those criteria had a pronounced effect on both scores and ranking of the third through fifth-ranked technologies. While both forms of SnCu solder remained the best-performers with relatively small changes in scores, Figure III-8 illustrates that the ordering and scores of tin/lead solder and the SAC forms of solder changed substantially.

Figure III-7
Garment Care: Missing Data Criteria Removal Under MAUT
Baseline Criteria Removal

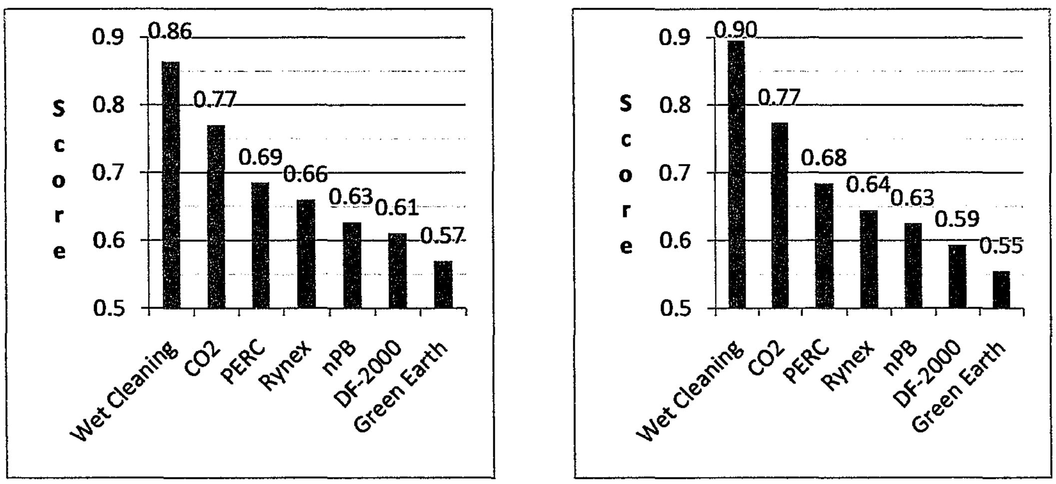
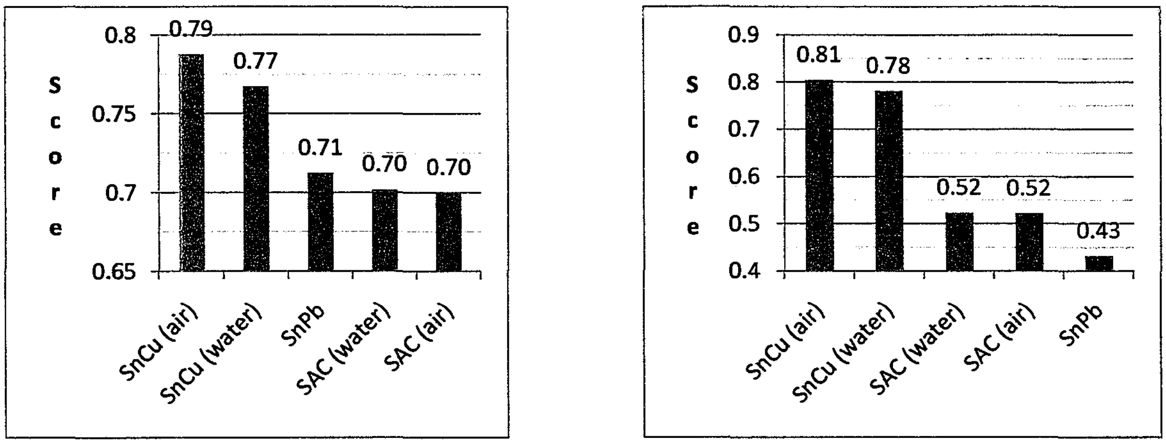


Figure III-8
Lead Solder: Missing Data Criteria Removal Under MAUT

Baseline | Criteria Removal

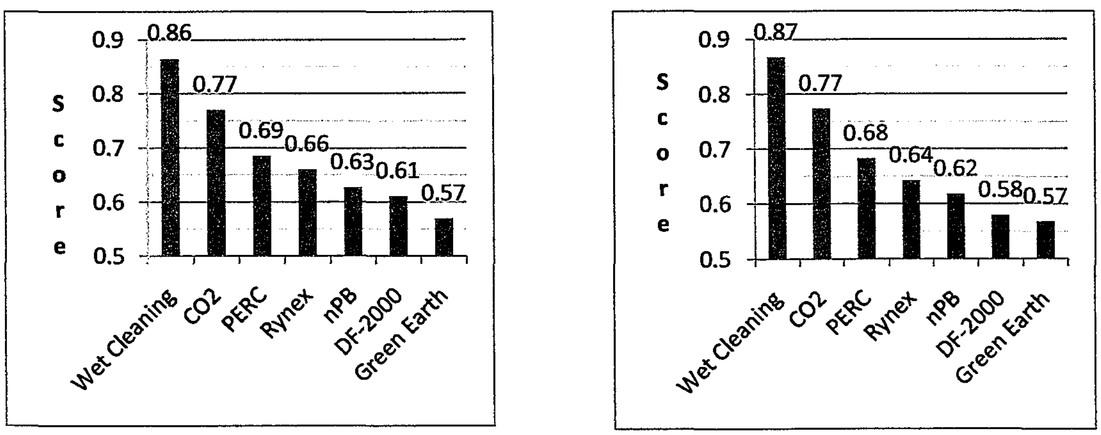


Under outranking, application of the criteria removal convention to the garment care case resulted in no change in ordering of the alternatives from the baseline ordering. For the lead solder case, criteria removal under outranking likewise resulted in no change in ordering of the alternatives.

Scenario 3: Incomplete Data. Unlike Scenario 1 (which assumed a mid-point for all missing and incomplete data), Scenario 3 instead makes a worst case assumption for all missing and incomplete data. In other words, Scenario 3 assumes that the alternative received the worst possible score for the criterion in question. As Figure III-9 shows, the adoption of these two different conventions had no impact on the ranking, and very little impact on the alternative scoring under MAUT.

Figure III-9
Garment Care: Missing Data Worst Case Assumption Under MAUT

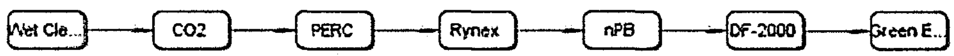
Baseline | Worst Case Assumption



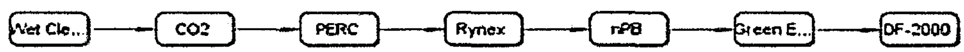
Under outranking, Figure III-10 shows that use of the worst case assumption resulted in a flip in the two lowest ranked alternatives as compared to the baseline (midpoint).

Figure III-10
Garment Care: Missing Data Worst Case Assumption Under PROMETHEE

Baseline



Worst Case Assumption



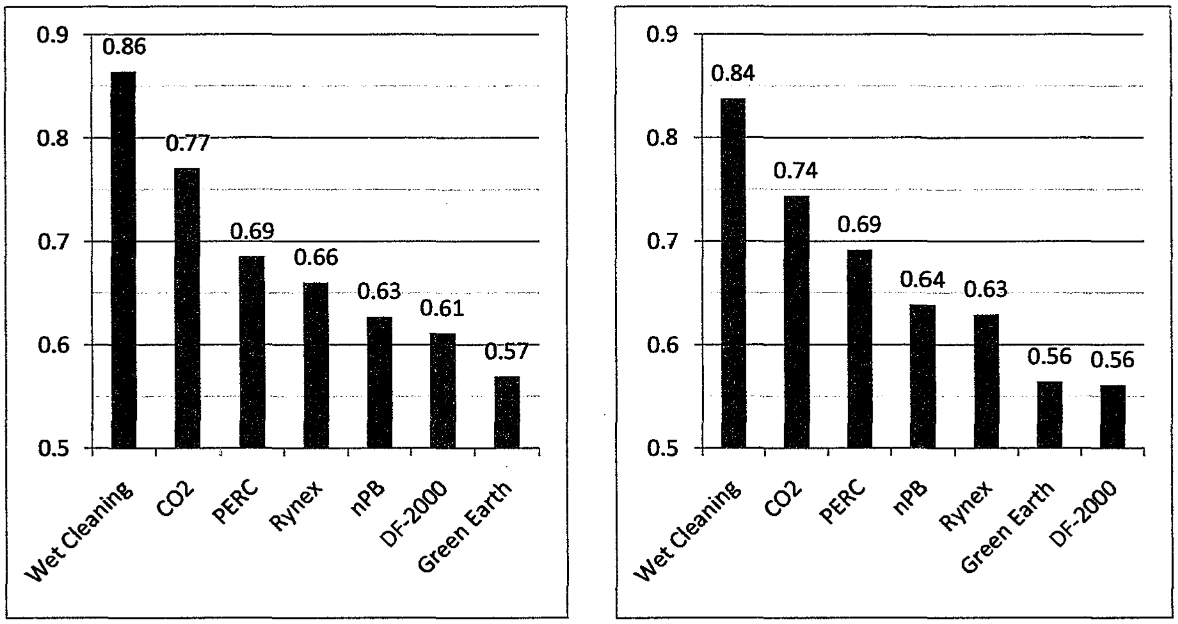
Scenario 4: Continuous vs. Categorical Data. Scenario 4 considers the impact of using categorical rather than continuous data in the garment care case. In some instances, categorical data may be the only data available, while in other cases some form of continuous data is available. In making comparisons between alternatives, continuous data will generally provide greater resolution; it will enable the decision-maker to draw finer distinctions between the alternatives. Some alternatives analysis methods adopt categorical approaches with respect to criteria for which continuous data is available. For example, under its Design for the Environment methodology, EPA groups continuous data regarding acute toxicity (in the form of median lethal dose or concentration (LD_{50} or LC_{50})) into four categories: Very High, High, Moderate, and Low.[60]

This scenario was designed to examine the impact of converting continuous data to categorical data in the case study. For this scenario, we used categories for flashpoint, acute toxicity, persistence, bioaccumulation, and aquatic toxicity. As Figure III-11 shows, the ordering under MAUT with respect to the top three performers (wet cleaning, CO_2 cleaning and perchloroethylene cleaning) is unchanged, but the remaining four alternatives have been significantly reordered.

[60] EPA, Design for the Environment Program Alternatives Assessment Criteria for Hazard Evaluation (Draft Jan. 2011)

Figure III-11
MAUT Outcomes Under Continuous and Categorical Approaches

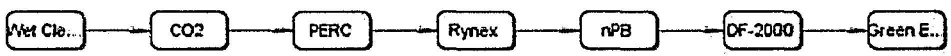
Scenario 1: Baseline (Continuous) **Scenario 4: Categorical**



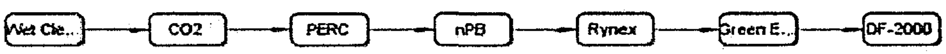
As Figure III-12 demonstrates, the same outcome occurs under outranking.

Figure III-12
PROMETHEE Outcomes Under Continuous and Categorical Approaches

Baseline (Continuous)



Scenario 4: Categorical



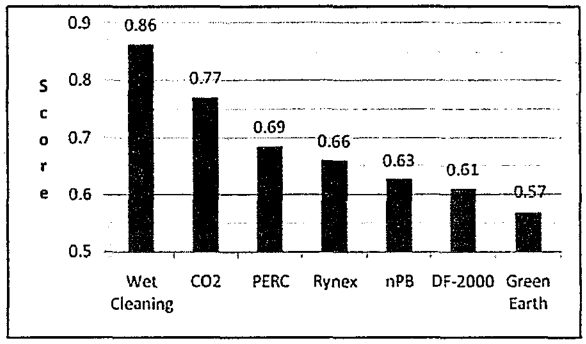
B.4 Effects of Stakeholders Weighting on Ranking

This set of scenarios considers the impact of adjustments to criteria weights. As noted above, in our baseline and other scenarios we used the average weights derived from the stakeholder elicitation process. In this set of scenarios, we systematically varied the weights to test the robustness of the outcomes under different stakeholder weighting regimes. Rather than using the average weighting from all individual stakeholders, in these scenarios we used the average weighting for each of the four stakeholder groups: Industry, Environmental NGO, Policymaker, and Consumer groups, respectively. The weighting scenarios were run on both the garment care case and the lead solder case.

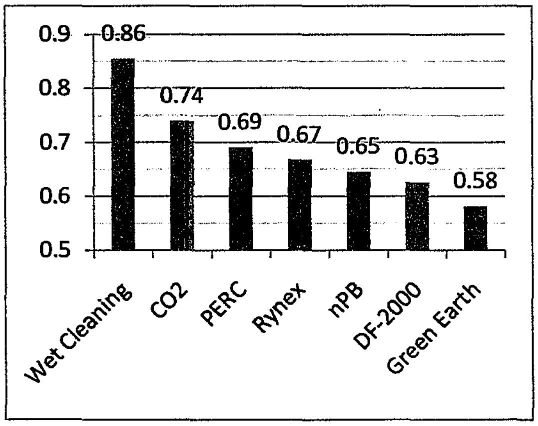
As Figure III-13 shows, there was very little variation in outcome for the garment care case attributable to the different weighting regimes in MAUT, with the scores slightly different and the order unchanged.

Figure III-13
Garment Care
Comparison of Results Under MAUT for Four Weighting Regimes

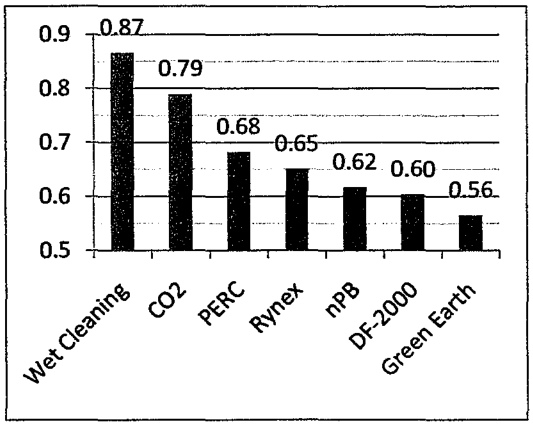
Baseline Weighting (Average)



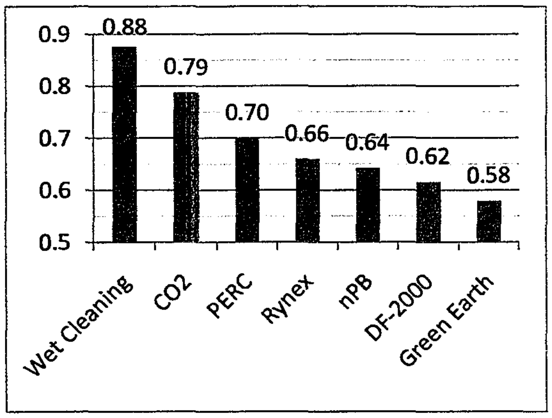
Industry Weighting



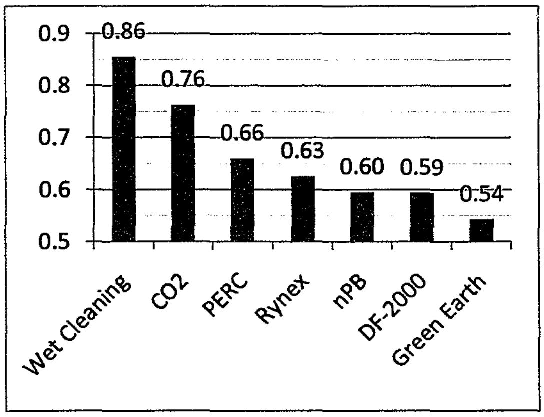
Environmental NGO Weighting



Policymaker Weighting



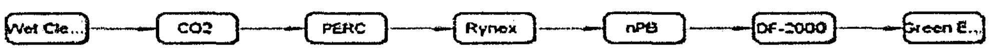
Consumer Weighting



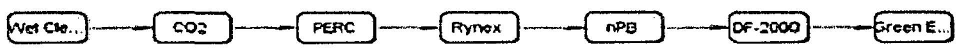
In outranking, Figure III-14 demonstrates that there was also little variation with the exception of nPB and DF-2000, which flipped between fifth and sixth place in order under the various scenarios. In PROMETHEE evaluations under each weighting scenario, those two technologies were very closely rated in terms of positive and negative flows; the relative shifts in the weighting across the regimes tipped the balance in different directions in the various scenarios.

Figure III-14
Comparison of Results Under PROMETHEE Four Weighting Regimes

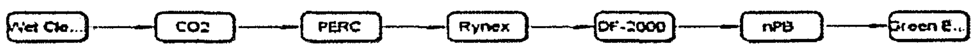
Baseline Weighting (Average)



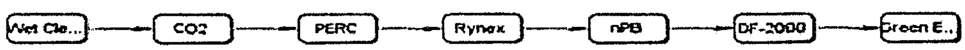
Industry Weighting



Environmental NGO Weighting



Consumer Weighting

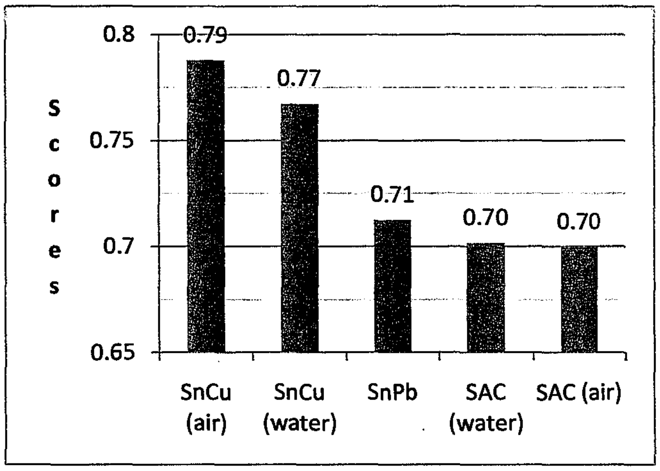


Policymaker Weighting

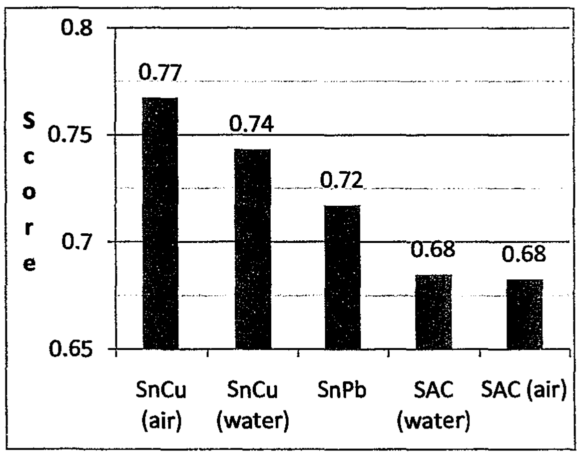
In the lead solder case, there was some small variation in scores across the different weighting regimes. The ordering of the alternatives was consistent across the regimes as well, with the exception of the policymaker weighting, which moved tin/lead solder from the third position to the last position in terms of overall performance. See Figure III-15. As Figure III-16 demonstrates, under outranking, none of the four weighting regimes affected the original ordering that resulted from the baseline scenario's use of average weights.

Figure III-15
Lead Solder
Comparison of Results Under MAUT for Four Weighting Regimes

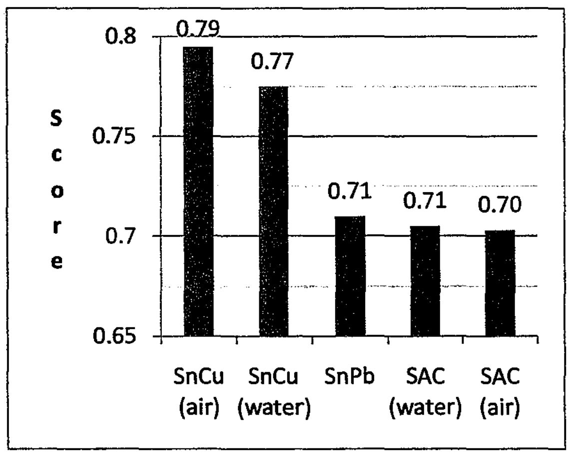
Baseline Weighting (Average)



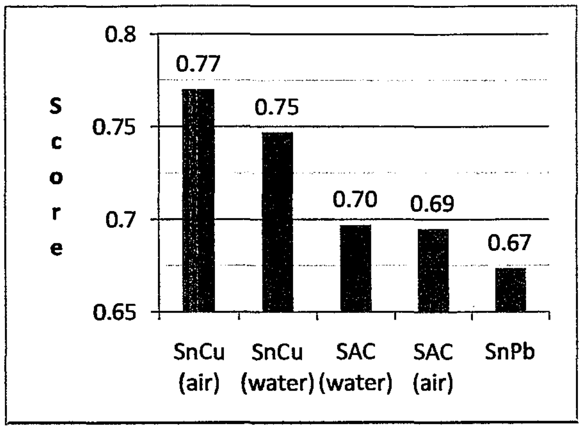
Industry Weighting



Environmental NGO Weighting



Policymaker Weighting



Consumer Weighting

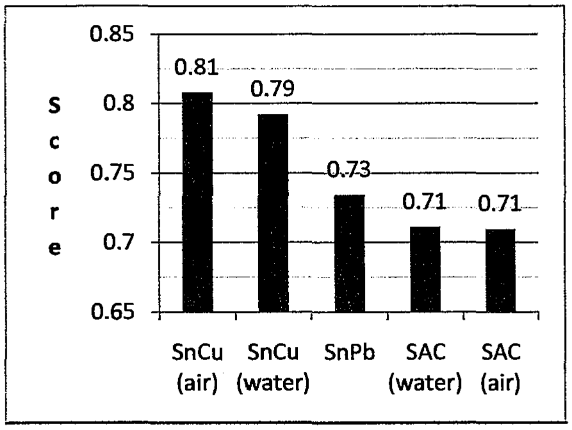
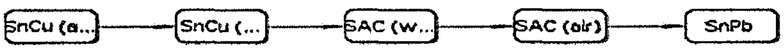
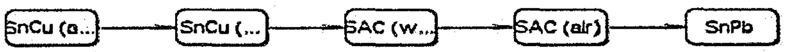


Figure III-16
Lead Solder
Comparison of Results Under PROMETHEE for Five Weighting Regimes

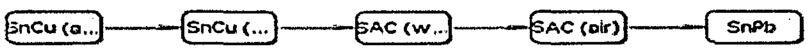
Baseline Weighting (Average)



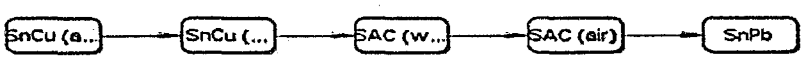
Industry Weighting



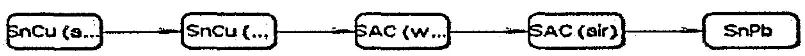
Environmental NGO Weighting



Consumer Weighting



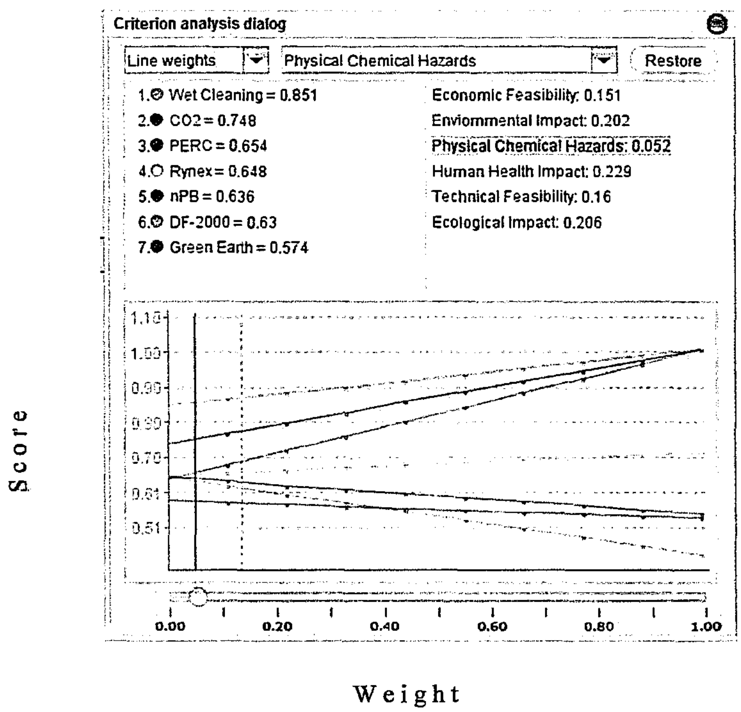
Policymaker Weighting



B.5 Sensitivity Analysis

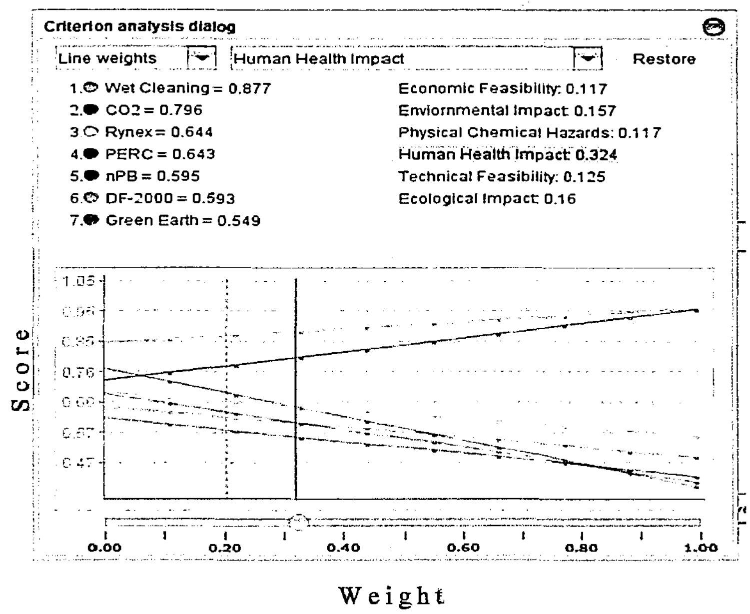
MCDA methods also allow for sensitivity analysis, a technique in which a variable is systematically modified to determine its impact on the outcome. In this case, we demonstrated this tool, using it to modify the weighting assigned to various criteria in both case studies. In the case of dry cleaning, we varied the weighting for physical/chemical hazards, the critical criterion on which perchloroethylene dry cleaning outperformed all alternatives save two. The baseline outcome is very robust in this context. As Figure III-17 indicates, the starting weight of 13.75% had to be reduced to just over 5% in order to allow just one more alternative—Rynex—to outperform perchloroethylene. (In Figure III-17, the dotted vertical line indicates the original weighting assigned to the physical/chemical hazard criteria. The solid vertical red line to its left indicates the adjusted percentage weight of 5.2%. The list of technologies in the upper left portion of the figure displays the rank order (from most preferred at the top to the least preferred at the bottom) with the adjusted weighting regime placing Rynex above perchloroethylene.)

Figure III-17
Garment Care Sensitivity Analysis: Physical/Chemical Hazard Weighting



The results with respect to human health impacts are similarly robust. As Figure III-18 shows, the weight assigned to human health impacts would have to be increased to more than 32%--an increase of 12 percentage points—in order to affect the outcome. At that point, as indicated at the intersection of Rynex and perchloroethylene at the solid vertical red line, Rynex's score rises above that of perchloroethylene.

Figure III-18
Garment Care Sensitivity Analysis: Human Health Impact Weighting



Both Figures III-17 and III-18 likewise demonstrate the robust nature of the ranking of wet cleaning and CO_2 cleaning as the top two alternatives. Those two alternatives are represented by the dark green and dark blue lines on the graphs. Wet cleaning retains the highest score without regard to the weightings of these two criteria, and CO_2 cleaning likewise remains in second position under most weightings.

The results of the lead solder case are less robust across different weightings. For this case study we modified the weighting for technical performance, a criterion on which both SAC forms of solder out-performed tin/lead solder. As Figure III-19 shows, at the baseline weighting of 14.5% for technical performance, tin/lead is ranked above the SAC solders considering all criteria. As the weighting placed on technical performance (and thus its importance to the outcome) is increased, lead's performance vis-à-vis SAC solders deteriorates. Indeed, as indicated by Figure III-20, tin/lead solder drops to the last position after the weighting for technical performance is increased by just over 4 percentage points to 18.7%. Finally, when the weighting of technical performance rises to 24.1%, SnCu (water quenched) is displaced from second position in the rank ordering by SAC (water quenched). See Figure III-21.

Figure III-19
Lead Solder Sensitivity Analysis: Original Technical Performance Weighting

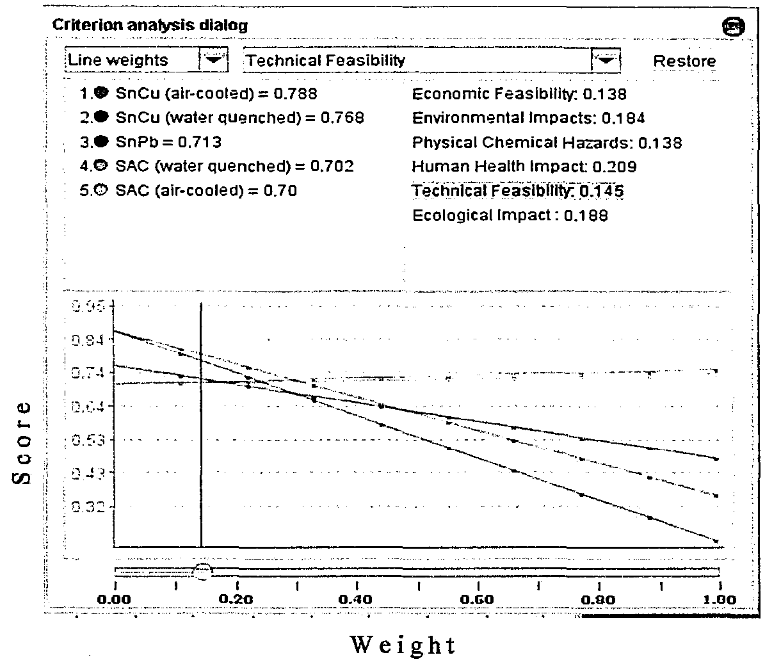


Figure III-20
Lead Solder Sensitivity Analysis: Modified Technical Performance Weighting (18.7%)

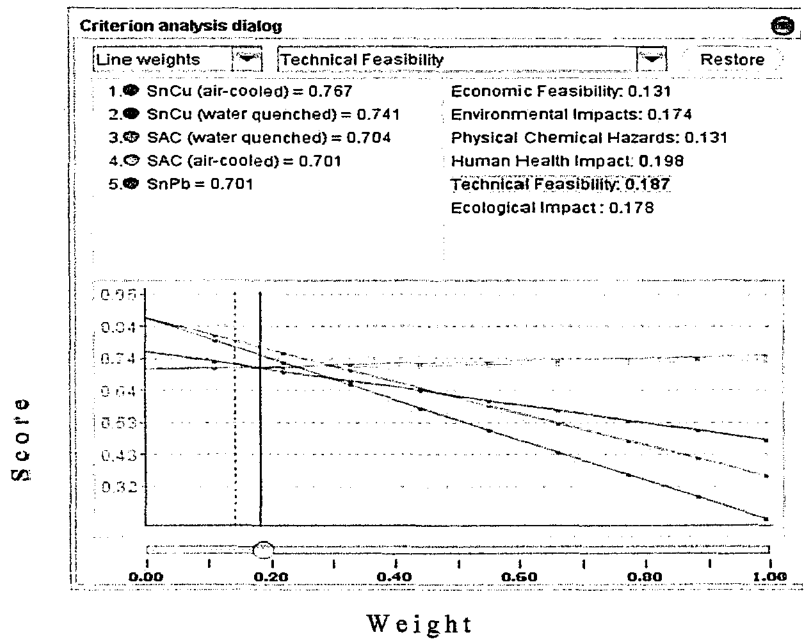
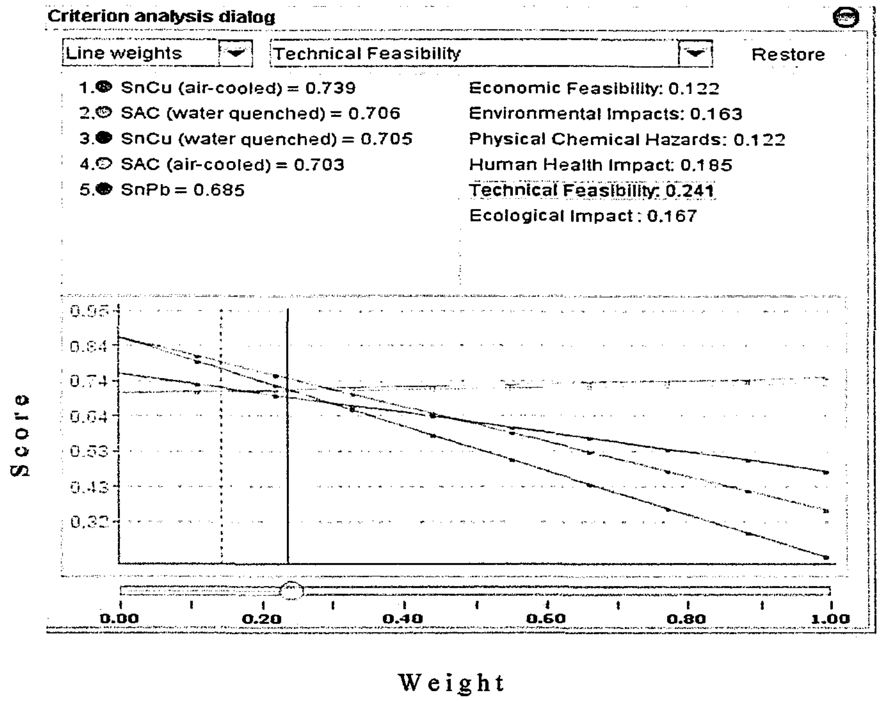


Figure III-21
Lead Solder Sensitivity Analysis: Modified Technical Performance Weighting (24.1%)



B.5 Decision Model Variations.

The final set of scenarios tests the robustness of the outcomes under different design parameters, two relating to the broader decision framework and one relating to specific modeling assumptions regarding utility functions under MAUT. With respect to the decision framework, our basic assumption was that even though the various criteria may have differing importance to the decision-maker, all criteria would be considered in comparing the alternatives. In other words, no single criterion would operate as a threshold factor to screen out alternatives from further consideration. The first two scenarios of this set alter that assumption, examining the outcomes if human health hazards or technical feasibility were established as threshold, screening factors. The remaining scenario in this set introduces a non-linear utility function designed to reflect potential decision-maker preferences more realistically. All scenarios in this set were run in the garment care case study only.

Table III-3
Model Design Modifications

Scenario Number	Scenario Name	Scenario Description
5	Sequential Decision Model	Using baseline, first run the model using just the Physical Chemical Hazards, Human Health, and Ecological criteria. Take top four

Scenario Number	Scenario Name	Scenario Description
		alternatives (including perchloroethylene if it is in the top four) and run the model again using just the Environmental criterion. Take top two alternatives (again including perchloroethylene if it is in the top two), and run the model using the Technical Feasibility criterion and the Economic Feasibility criterion.
6	Modified Sequential Model	Same as Scenario 5, but reversing the sequence to start with evaluation under the Technical Feasibility and the Economic Feasibility criteria.
7	Utility Functions	Evaluating the alternatives on the basis of Physical Chemical Hazards, Human Health, and Ecological criteria, but inserting utility functions for **inhalation LC50, oral LD50, and dermal LD50.**

Scenario 5: Sequential Decision Model. Table III-4 shows the results from the Sequential Decision Model. In the first part of the sequential model under MAUT, the top four performing technologies regarding Physical Chemical Hazards criterion, Human Health criterion, and Ecological criterion were wet cleaning, CO_2 cleaning, perchloroethylene cleaning, and Rynex. Those four were carried on for further evaluation under the Environmental criterion, resulting in identification of wet cleaning and CO_2 cleaning as the top two remaining performers. Further evaluation under the Technical Feasibility criterion and the Economic Feasibility criterion resulted in identification of wet cleaning as the best performing alternative.

Table III-4
Sequential Decision Model Results

First Screen: Top 4 after Physical/Chemical, Human Health, and Ecological	**Second Screen: Top 2 after Environmental Impact**	**Third Screen: Technical Performance and Economic Feasibility**
Wet Cleaning	Wet Cleaning	Wet Cleaning
CO2	CO2	
Perchloroethylene		
Rynex		

Scenario 6: Modified Sequential Decision Model. Table III-5 shows the results from the Modified Sequential Decision Model. The modified sequential model examines the outcome in the event that Technical and Economic Feasibility were used as an initial screen, followed by

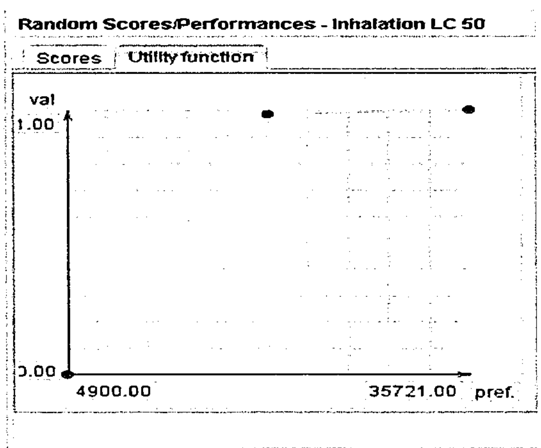
consideration of the Physical Chemical Hazards, Human Health, and Ecological criteria, and then by the Environmental criteria. The top four performing technologies concerning Technical and Economic Feasibility were perchloroethylene cleaning, nPB, wet cleaning and Rynex in that order. Those four were further evaluated under Physical Chemical Hazards, Human Health, and Ecological criteria, resulting in two top performers: wet cleaning and perchloroethylene cleaning. The final evaluation again resulted in the top ranking for wet cleaning.

Table III-5
Modified Sequential Decision Model Results

First Screen: Top 4 after Technical Performance and Economic Feasibility	Second Screen: Top 2 after Physical/Chemical, Human Health, and Ecological	Third Screen: Environmental Impact
Perc	Wet Cleaning	Wet Cleaning
nPB	Perc	
Wet Cleaning		
Rynex		

Scenario 7: Utility Functions. The last scenario in this set modifies the default linear utility function to demonstrate the capacity of MAUT to account for different types of preferences. We evaluated the garment care alternatives with respect to Physical Chemical Hazards criteria, Human Health criteria, and Ecological criteria as above, with one change. In this scenario, we modified the linear utility function for acute toxicity (as measured by the median lethal dose or concentration (LD_{50} or LC_{50}) for oral, dermal and inhalation exposures.) Here we assumed that once the dose/concentrations reached a relatively high level (i.e., a "safe" level), decision-makers would be indifferent to further increases in the level. Figure III-23 shows the shape of the utility function:

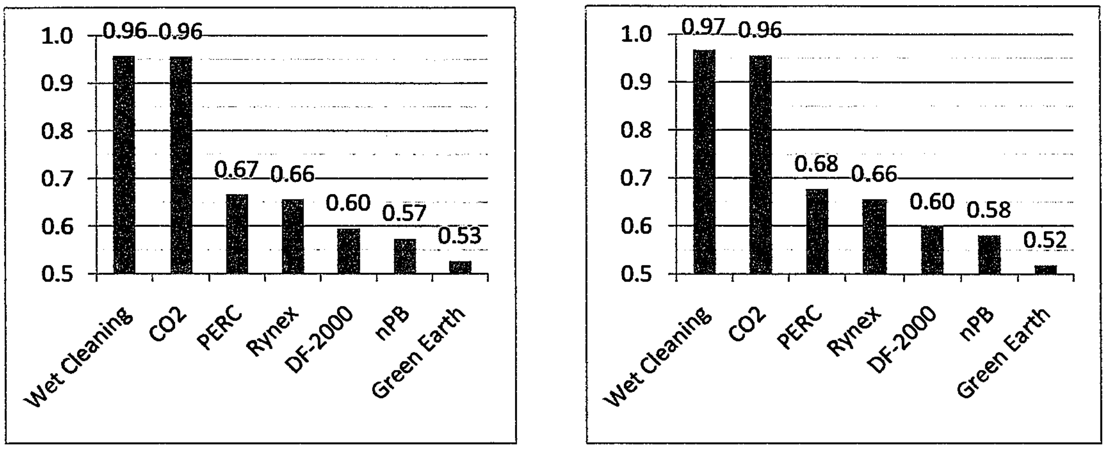
Figure III-23
Utility Function: Acute Toxicity



As Figure III-24 demonstrates, use of the utility function resulted in no significant change to scores or ordering under MAUT.

Figure III-24
Comparison of Linear and Modified Utility Function

Linear Utility Function **Modified Utility Function**



IV. General Conclusions and Recommendations

This project had two primary goals with respect to alternatives analysis: (i) explore the development of a rigorous alternatives assessment methodology that is consistent with the mandates of AB1879, and (ii) examine whether MCDA methods (including weighting approaches) may be appropriate for alternatives evaluation. The specific approaches and methods developed and applied as part of the project were not intended to be applied directly to alternatives analysis under AB 1879; rather they were designed to inform and enhance the development of regulatory alternatives analysis generally. This section summarizes the lessons learned from the project, and identifies several areas for additional work. This feasibility study developed a workable, comprehensive alternatives analysis model, and demonstrates the promise of multi-criteria decision analysis (MCDA) as a robust method to assist in alternative analysis. That said, a number of steps are needed to take this from a feasibility stage to a generally applicable methodology that is rigorous enough to provide consistent results but flexible enough to adapt to the variety of products-of-concern likely to be regulated.

A. Alternatives Assessment

Regarding alternatives assessment, the project provides a fairly extensive set of criteria, ranging from general upper level criteria to very specific measurement sub-criteria. The breadth of the criteria provides for a comprehensive evaluation of alternatives, significantly broader in reach that the set of indicators typically used in life cycle impact assessment. The project criteria establish generalized but measurable sub-criteria for technical and economic feasibility, a significant enhancement to existing alternatives analysis methods currently available. However, the comprehensive nature of the criteria creates commensurately greater data collection and management efforts. The amount of missing and incomplete data in the project's performance matrices, particularly in the lead solder case study, highlights the need for meaningful data generation and collection elements in the regulatory program. (Absent more data, the project demonstrates several approaches for dealing with missing and incomplete data, discussed in more detail below.)

The generic alternatives assessment model presented here requires further refinement is a number of areas. Perhaps most importantly, the model should be modified to capture the potential impacts of alternatives more completely. One clear example of this is the fairly positive ranking achieved by perchloroethylene drycleaning. As designed, the model failed to reflect the pervasive groundwater contamination stemming from perchloroethylene releases from drycleaning. While the model captures the relative carcinogenicity of perchloroethylene, it lacked an effective way of integrating the exposure issues associated with groundwater releases (although much of that may be a legacy issue). Additionally, it does not address the broader costs of that contamination to society, due to our decision to exclude societal level economic impacts due to the assumed "permit-based" approach. Other additional work includes:

- Development of specific principles for determining whether a criterion should be placed as a high level criterion or a lower level sub-criterion. Section II.C of the report discusses how the generic model used in the study was developed, including the placement of criteria at higher or lower levels. Such decisions were based on the judgment and experience of the project team, but a more systematic approach should be utilized for regulatory purposes.

- Creation of clear standards for dealing with overlaps between criteria. Section II.C discussed specific cases in which a measurement sub-criterion was removed because it overlapped with another existing criterion. Generally speaking, we allowed overlap where the same attribute in question exhibited distinct impacts in different areas. For example, energy use associated with an alternative could have both energy conservation impacts (primarily a societal concern) and energy cost impacts (primarily an individual facility concern). Conversely, where the two impacts associated with an attribute were directly linked, we selected one measurement sub-criteria to capture those impacts. Take the case of chemicals that are listed as hazardous air pollutants. One could reach this through both a measurement sub-criterion under air quality, and a measurement sub-criterion under human health impacts. In that case, we chose human health impacts only so as to avoid double-counting. The general decision rule we used here should be more fully articulated and tested in other circumstances.

- Accounting for interaction of variables. The assumption used in the generic model was that criteria and sub-criteria within the same level and at different levels were independent of one another. Further work is required to test this assumption across all categories, and to refine the model in cases in which there is some interaction. The hazard and exposure criteria highlight this concern. For many stakeholders, assessment of overall impact requires consideration of the specific hazard and the related exposure together, a perspective that our model and others do not fully capture.

- Accounting for the quality of data. The generic model treats all data as essentially of the same quality. Further work is needed to integrate relative data quality into the model.

B. Alternatives Evaluation

The project results demonstrate the viability of MCDA methods to assist in the evaluation of complex alternatives assessment data. In particular, the case studies show that the MCDA models can provide decision-makers and stakeholders with a transparent evaluation of that data, presenting a ranking of alternatives accompanied with explanations of how the alternatives' performance on various criteria affected that ordering. The methods also allow parties to understand how their weighting affects outcomes. The methods also permit decision-makers to adjust the MCDA method's assumptions regarding the nature of their preferences. For example, by altering the shape of utility functions, MAUT can capture situations in which a user is less concerned about a criterion where performance on that criterion is above or below certain ranges. Outranking offers similar flexibility.

The project results provided useful insights on several aspects of alternatives evaluation. First, the stakeholder weighting process revealed that there was relatively little difference across groups with respect to the relative weight they placed on the criteria. This was driven home by the lack of substantial differences in the ranking of alternatives in either case study by the four groups. Of course, the elicitation process was limited in terms of sample size; nonetheless this outcome raises interesting questions regarding just how much disagreement there will be regarding weighting among the groups as the regulatory process moves forward.

Second, the results regarding missing data demonstrate that how a method handles missing data can significantly affect the outcome of the evaluation. In the lead solder case, there was a substantial amount of missing data, and the ranking of alternatives differed significantly depending upon whether we used mid-points to fill in the missing data or simply removed those criteria for which there was no data. Conceptually, it is clear that inserting default data for all alternatives (such as a mid-point or worst-case value) will dilute the impact of other criteria for which there *is* data as compared to removing the criteria for which data is missing. For compensatory methods in which good performance on one criterion can offset bad performance on another, inserting identical default values for all alternatives would have an equalizing effect. Accordingly, it is not clear that inserting a default value—even a very "conservative" worst case value—would necessarily have an overall protective effect.[61] However, in cases in which certain criteria are simply not relevant to the alternatives, it would be appropriate to remove those criteria from the model at the outset rather than filling them with uniform default values for all alternatives.

Third, the project outcomes indicate that the two MCDA approaches are fairly robust, particularly in identifying the top ranked alternatives across a range of scenarios that shifted weighting, data assumptions and model design. The results of the weighting sensitivity analysis in the lead solder case offers a cautionary note, however, demonstrating that, depending upon the

[61] This dilemma emphasizes the importance of having meaningful, effective data generation and submission elements in the regulations.

specifics of a given case, shifts in how criteria are weighted can have significant impacts on relative ranking.

As with alternatives assessment, the alternatives evaluation component also requires additional development in a variety of areas:

- Further evaluation of MCDA Methods. While the project focused upon two leading MCDA methods, other methods with additional useful features are available. Moreover, the project identified some differences in outcomes as between MAUT and outranking. For example, while the removal of all criteria with missing data in the lead solder case study resulted in significant reordering of alternatives ranking in MAUT, in outranking it had little effect. Additional research and case studies should be performed to explore those differences, and to develop standards for selecting the most appropriate MCDA method for various application in alternatives analysis.

- Explore alternative decision frameworks. This project assumed that the regulatory alternative analysis would adopt a compensatory approach in which all criteria are considered together. We modified this slightly in Scenarios 5 and 6 in Sequential Model and the Modified Sequential Model. Policymakers may instead adopt a different framework. For example, under "goal aspiration" or threshold decision frameworks, the decision-maker does not seek to identify the best option or even rank the options. Rather, such methods identify those options that achieve (or come closest to achieving) some minimum level of performance for one or more criteria.[62] Further research could examine alternative frameworks and relevant appropriate MCDA methods.

- Further development of weighting for regulatory purposes. The project engaged in limited stakeholder elicitation and weighting. Further research regarding alternative weighting methodologies and the integration of weighting into the regulatory framework would be useful.

---

[62] Igor Linkov, *et. al*, *Multi-Criteria Decision Analysis: A Framework for Structuring Remedial Decisions at Contaminated Site*, in COMPARATIVE RISK ASSESSMENT AND ENVIRONMENTAL DECISION MAKING 15, 17-20 (Igor Linkov and A. Bakr Ramadan eds. 2004).